

# Beware of Test Scores Masquerading as Data

John Merrifield  
February 28, 2015

A semi-taboo area of insufficient discussion is the reliability of the test score data from the statewide, nationwide, and international standard tests; for example, our National Assessment of Educational Progress (NAEP), but not nearly just the NAEP test scores. You can learn about all of the reliability issues from experts like [Richard Phelps](#), and [Richard Innes](#).

I have frequently raised concerns about test score data generated by exams that don't impact the students that take them; that is, where a poor effort by a student does not adversely impact the student. The norm for most international, national, and some statewide standardized testing is that the students taking them have no tangible incentive to give their top effort. NAEP – the so-called nation's report card is among the no-stakes-for-the-students. Expressing a concern for that data reliability issue in an e-mail or a conversation nearly always yields no response, or a vague, dismissive response; something approaching '[emperor has no clothes](#)' proportions.

The discovery that prompted this blog was Richard Phelps's [pronouncement](#) that:

“Indeed, one drawback to the standardized student tests with no stakes for the students is that student effort does not just vary, it varies differently by demographic sub-group. The economists who like to use such scores for measuring school and teacher value-added just assume away these and other critical flaws.”

So, while such test scores might be broadly accurate – more substantive persuasion please – they may just be numbers masquerading as data for some of the uses they have been put to. And it's another reason to question the current system's extensive reliance on top-down-only-accountability to formal authority that must be based on objective apple-to-apple comparisons. It's a narrow basis for accountability, and vulnerable to distortion, including fraudulently. We need robust universal parental school choice to exploit subjective, bottom-up-accountability to clients; the dominant form

of accountability in our economy. To orchestrate a high level of performance from diverse children and educators, we need a system that employs a mix of top-down and bottom-up accountability.

I'm willing to rely on NAEP and [PISA](#) test score data (etc.), with some reservations and reticence, because the data are consistent with the high stakes-generated data and other indicators of school system effectiveness, and with established economic theory. But the no-stakes-for-the-students test score issue needs a lot more study and discussion. Denial is more than a river in Egypt.