



Is the Smarter Balanced National Test Broken?

By [Douglas J. McRae](#) & [Williamson M. Evers](#)

January 04, 2018

Across the country, [14 states](#) used the federally-funded Smarter Balanced tests as part of their statewide K-12 testing programs in spring 2017. But the results for this year have an integrity problem.

The contrast between the 2016 and 2017 scores for Smarter Balanced states is stark. Gains dropped like a stone. The 2017 results show 13 states with either zero gains or declines and only one state showing a tiny positive gain.

In contrast, the 2016 gains were typical annual gains – an average of 2.46 percentage points. With only one exception, all states had positive gains ranging from 1.80 to 3.75 percentage points.

One of us (McRae) collected overall “Percent Proficient and Above” data for each state using these national tests since 2015, and [computed overall gains](#) using a metric that permits interpretation like a 4-point Grade Point Average, with 3 and 4 point gains being good and very good and typical year-to-year gains being roughly 2 percentage points. This metric was based on [analysis](#) provided by the late Robert L. Linn, former president of the American Educational Research Association (AERA) and of the National Council on Measurement in Education (NCME), in the early 2000s.

Declines in annual statewide assessments are rare for tests like these, which are designed to provide comparable scores from year to year. A decline stands out like a sore thumb. For example, using the same metric, McRae [calculated](#) GPA-like gains for California’s previous statewide test from 2002 to 2013, and California recorded positive gains for 11 of the 12 years. There was only one year in which scores declined, but it was the last year, when the tests were not taken as seriously in anticipation of the switch to the then-new Smarter Balanced.

It is quite unlikely that the 2017 Smarter Balanced declines, based on roughly [six million students tested](#), reflect actual performance declines for students and schools across all those states. This is, after all, the third year of this national testing program. Those who watch testing results would expect slight gains, not a drop or a plateau.

Drilling into the data for 2017, it is apparent that the 2017 Smarter Balanced results are much worse for English

Language Arts (ELA) than for Mathematics. For ELA, 12 of 13 Smarter Balanced states had declines averaging -1.48 percentage points while a single state – California – had a miniscule gain of +0.10 percentage point. For Mathematics, seven states had positive gains and six states had either no gains or losses.

Smarter Balanced is stonewalling efforts to figure out what has occurred. It **refuses to acknowledge** that the 2017 scores are highly unusual and, instead, claims the scores are just normal year-to-year fluctuations of gain scores. That argument is hogwash. It is totally inconsistent with the actual 2017 consortium-wide gain data.

Others agree with us. For example, **Ed Haertel**, a Stanford University professor and a member of the Smarter Balanced Technical Advisory Committee, told EdSource.org, “These are not merely random-chance fluctuations.” He’s doubtful that “there was some slight overall decline in students’ overall proficiency.”

What needs to be done? Smarter Balanced managers need to open the wall of secrecy surrounding the technical information needed to investigate this gain score problem. We suggest that information on changes in the large test question banks used from 2016 to 2017, as well as adjustments to the scoring formulas for 2017, are areas to investigate first. None of this information involves unveiling test questions themselves – the usual reason cited for having test analyses conducted behind closed doors. Any conclusions should be verified by independent experts. This has to be done within weeks, not months, with results provided immediately for all users of Smarter Balanced 2017 scores. What happened? We need to know.

We believe that some problem caused the 2016 and 2017 Smarter Balanced scores not to be comparable, creating gain scores that are not accurate reflections of achievement trends, at least for ELA gain scores. The consortium-wide gain score of -1.48 percentage points for ELA just doesn’t make sense given that year-to-year gains are typically in the +1.00 to +3.00 range.

What is affected? It could be scores and perceived gains or declines at the individual, school, district, and state levels.

If Smarter Balanced wants to have credibility for its analysis of what happened, it should allow independent researchers access to the appropriate technical data. Right now, Smarter Balanced is characterized by a lack of transparency. The process is opaque. American Institutes of Research is looking into the 2017 gain-score issue but is a major subcontractor for Smarter Balanced. It can hardly be an unbiased source for an independent review of work done by Smarter Balanced itself or its contractors.

Smarter Balanced has to find the problem and issue corrected scores, just as an automobile manufacturer has to correct a problem with exploding airbags. An incorrect-scores problem for approximately six million kids over 14 states is a big problem.

If Smarter Balanced needs to withdraw its previously announced scores and issue new ones, do it. If the consortium needs to withdraw the entire test and nullify the scores on it, do it. The integrity of statewide tests and scores depend on Smarter Balanced doing the right thing.

Douglas J. McRae is a retired educational measurement specialist from Monterey, California. Williamson M. Evers is a research fellow at Stanford University’s Hoover Institution and a former U.S. Assistant Secretary of Education for Planning, Evaluation and Policy Development.