

RAND versus RAND

What Do Test Scores in Texas Tell Us? by Stephen P. Klein et al.

By [Eric A. Hanushek](#)

|||

Spring 2001 / Vol. 1, No. 1



The Sequel

What Do Test Scores in Texas Tell Us?

by Stephen P. Klein, Laura S. Hamilton, Daniel F. McCaffrey, and Brian M. Stecher

RAND Corporation, 2000.

Just two weeks before the presidential election, yet another team of RAND researchers released a short paper that seemingly contradicted Grissmer et al.'s celebration of Texas's achievement gains on the NAEP. RAND II found only small NAEP achievement gains in Texas, similar to those nationwide and contrasting sharply with "soaring" scores on the Texas Assessment of Academic Skills (TAAS). These disparities, the authors suggested, point to potentially serious flaws in Texas's state-run testing program.

The direct conflicts of RAND I and RAND II underscore the fact that RAND is a collection of franchisees. The parent company attempts to maintain some degree of quality control but ultimately is not able fully to adjudicate quality—particularly, one suspects, when the answers are fuzzy and when the sponsor pressures are high.

RAND II presents two separate analyses that, taken together, seem to undermine Texas students' spectacular gains on the TAAS. First, Texas students showed substantially more improvement on the TAAS than they did on the NAEP during the 1990s. Second, in a sample of 20 schools that the authors had collected for other purposes, the expected negative relationship between a student's TAAS score and his eligibility for the federal school lunch program, a common measure of disadvantage, didn't arise on the TAAS. This latter finding led RAND II to conclude not just that the TAAS is a poor instrument but also that high-stakes testing leads to the artificial inflating of scores through "teaching to the test," especially for disadvantaged students.

It should not be particularly surprising that student performance improved more dramatically on a test that was aligned with a particular state's curriculum (the TAAS) than on a more generic test of subject matter (the NAEP). Thus, while the question of the TAAS test's validity is an important one, the simple evidence presented in RAND II falls very short of yielding any solid answers.

Likewise, the fact that data on 20 schools show a peculiar relationship with any variable is unremarkable. After all, even if the authors attempted to draw a representative sample—which they did not—the idiosyncrasies of such a small sample would preclude any ability to generalize. Indeed, a simple plot or a formal statistical analysis of TAAS scores across all Texas schools reveals a clear, and expected, strong negative relationship between students' scores and their eligibility for subsidized school lunches.

The point of clearest conflict with RAND I is the consideration of NAEP performance. RAND I—not as focused on the relationship between its statistics and presidential campaigns—considered all seven NAEP tests

given between 1990 and 1996 and attempted to adjust for differences in the students' backgrounds. The result was high marks for Texas's performance improvements on the NAEP. RAND II, by contrast, ignored student background, placed more weight on a different subset of test results (including the 1998 results, which were not included in RAND I), used somewhat different approaches, and concluded that there was nothing special about performance in Texas.

What lessons might we take away from the RAND I vs. RAND II debate?

- Analyses of small amounts of imperfect data can yield widely different conclusions. Such analyses should be heavily discounted.
- Consideration of a study's quality tends to get lost in the ensuing policy discussion. Neither RAND study holds up to a modicum of scrutiny.
- The desire for publicity apparently pushes some researchers to prepackage their own sound bites. The PR blitzes that accompanied both RAND I and RAND II undermined any public discussion of what turns out to be relatively impotent research designs.
- Journalists tend to judge a study's quality—particularly a complicated statistical study—by its conclusions and by an undue emphasis on the study's source rather than the strength of its analysis. RAND's undeniable history of producing solid research doesn't mean that every study under the RAND imprimatur deserves unquestioned repeating.

The result is a distorted and unhealthy policy discussion.

0
tweets

[User Agreement](#) | [Privacy Policy](#)

[Reporting Copyright Infringement](#) | [Guidelines for Submissions](#) | [Permissions](#) | [FAQ](#)

Web-only content Copyright © 2011 President & Fellows of Harvard College
Journal content Copyright © 2011 by the Board of Trustees of Leland Stanford Junior University

Business Office

Program on Education Policy and Governance

Harvard Kennedy School

79 JFK Street, Cambridge, MA 02138

Phone (877) 476-5354 Fax (617) 496-1507