

PART 6 CHAPTER 7: As the Federal Court Case Was in Process with Looming Decision, Dallas I.S.D. and the Prestigious Rand Corporation Were Peeing on the Parade & So Was My Small Non-Profit in Houston

This chapter is the beginning of the dismemberment of any pretense that the initial TAAS testing program was troubled by even the notion of academic integrity. In its November 1998 letter to Texas Education Commissioner Dr. Moses, Dallas I.S.D. provided the TEA the results of its study which proved with statistical certitude:

- Passing the TAAS’ **reading** tests correlated to the 10th to 27th percentiles in performance on the 3rd Grade through 8th Grade and the 10th Grade high school exit test on the national Iowa Test of Basic Skills (ITBS).
 - At the high school exit level of 10th grade, the statistical correlation was the 10th percentile.
 - At the third-grade level, the statistical correlation was the 22nd percentile.

Approximate Percentiles Correlating TAAS Passing Performance With National Percentile On ITBS

Reported By Dallas ISD on TAAS 1998 Tests

MATH		READING	
Grade	% Tile	Grade	% Tile
3rd	40	3rd	22
4th	42	4th	27
5th	40	5th	26
6th	33	6th	26
7th	33	7th	24
8th	31	8th	22
10th	23	10th	10

Approximate Percentiles Correlating TAAS Scores of 80% Content Mastery or Higher With ITBS Percentiles

MATH		READING	
Grade	% Tile	Grade	% Tile
3rd	62	3rd	36
4th	66	4th	41
5th	57	5th	36
6th	55	6th	35
7th	55	7th	36
8th	53	8th	33
10th	38	10th	16

Approximate Percentiles Correlating TAAS Scores of 90% Content Mastery or Higher With ITBS Percentiles

MATH		READING	
Grade	% Tile	Grade	% Tile
3rd	96	3rd	60
4th	97	4th	67
5th	87	5th	56
6th	88	6th	60
7th	93	7th	64
8th	90	8th	60
10th	84	10th	53

- Passing the TAAS’ **math** tests correlated to the 23rd to 42nd percentile on the ITBS.
 - At the high school exit level of 10th grade, the statistical correlation was the 23rd percentile
 - At the third-grade level, the statistical correlation was the 40th percentile.

Here’s the bottom line. Dallas ISD confirmed that the passing standard of TAAS overall correlated to extremely deficient, below grade-level standards on a national normed-referenced test. The Dallas study involved right at 60,000 students district wide who took both the TAAS reading and math tests and the reading and math assessments in the ITBS. Passing TAAS on a nationally recognized independent grade-level assessment was substandard academically.

We have a due-diligence obligation to note that the Dallas communication included its representation that performance at the high end of the TAAS tests was found to have measured “high level performance.”

In reality, Dallas ISD misread and misinterpreted its own data at the higher levels of TAAS performance because it did not take into consideration a so-called topping out factor in these circumstances that other statisticians will address later in this report.

Stated simply, what Dallas ISD data shows is that TAAS was so far below grade-level that poor and average students could perform at the higher performance standards of each test but STILL BE well below a genuine, academically honest grade level for that grade and subject.

This topping out factor skews the correlation dramatically at the upper levels of measurement but not at the lower levels such as passing.

The full report is available. However, basic honesty mandates that we include their representations in the context of our analysis.

It was statisticians who have worked for the Rand Corporation nationally and at St. Thomas University in Houston who carefully explained this quirk in the Dallas ISD report as it related to high level academic criteria.

As a matter of history, the TEA's strong response to the district and the failure of Texas news media education reporters to even know about or understand the gravity of the hard data in the report basically kept it under wraps.

Not so as it relates to Dr. Stephen Klein, a lead education researcher for the Rand Corporation, who made a presentation at that same June 1999 NAS national conference at which Texas and the TEA's Smisco praised the integrity of TAAS. Dr. Klein's presentation created shock waves back in Texas' political establishment with a governor running for President. It did not go over well. Others had raised warning flags before this NAS conference. They were well-documented; extremely credible; driven by empirical data.

None were offered publicly that bore the name THE RAND CORPORATION. With a governor running for President of the United States and a federal court decision on the horizon, Rand brought gravitas to the debate, and that's why this chain of events is so critical in linking the original era of Texas accountability to the current era.

His oral review of Texas included zingers such as:

- **“...I am not saying that these people cheated for anything like that. I know there's something wrong. It's not right. There's nobody here who would say that there's not a strong correlation between socio-economic status and test scores. We see it in everything...”**

Dr. Klein advised the NAS conference that the Rand Corporation was conducting a national education project at 11 sites throughout the United States. It was a joint project focusing upon math and science with tests developed by Stanford and the Rand Corporation. Dr. Klein acknowledged the tests were not parallel to all participating districts or states around the country.

The plan was to administer its tests to students around the country who had taken that state's own test(s). The goal was to look at correlation analysis particularly involving socio-economic factors. At no time did Dr. Klein represent his conclusions as the result of a formal, extensive, Rand-quality full statistical analysis.

The testing movement was growing nationally; Rand began an initial project around the country; had a particular interest in Texas because of the national acclaim Texas was receiving due to dramatic closure of academic achievement gaps for at-risk, disadvantaged students.

Texas' reaction to Dr. Klein's analysis forced Rand to produce such a study proving once again that the old phrase “be careful what you ask for because you might get” never proved more predictive.

First, let's look at key excerpts from Dr. Klein's presentation referencing the Rand's insights into the TAAS test. The full report is available. He acknowledged that because of the asserted success of the TAAS testing program, the Rand was considering use of the TAAS test as a usable standard in national studies.

- “...Then we took a look at the correlation between those same measures and the TAAS, and it blew up. Looks like somebody had hit this thing with a shotgun. Free and reduced lunch. Here is the mean math. And the correlation is a .04. No relationship at all. Same kids. Exactly the same kids two weeks later. We had individual scores...”

- “...What happens when that correlation goes away and then two weeks later it pops back up again with those same students? It’s not as if they learned something and then forgot it. That doesn’t happen. And it’s not like the level came up or something like that. That’s what unique about these data. It’s exactly the same kids one-for one. How could they suddenly do so poorly? So that’s why I am suspect about the scores. So that’s why I’m suspect about the scores. It’s not a case of somebody coming and saying these kids really did excel, they really did learn a lot, and so on and so forth. But they did coming in with an alternative test should produce those results right back again? It didn’t happen. That the part that concerns us...
- “...Now. This is not an outcome that we wanted to find at all because this poses real problems for us because we had hoped to use TAAS scores. I don’t feel comfortable doing that anymore given these results because I think the scores are suspect. There’s a lot of possible explanations for what happened. Not one. But many possible explanations for what occurred here...”

The force of his conclusions and how he expressed them plus the fact that the Rand used alternative criterion tests prompted strong reactive questions for defenders of the Texas’ asserted progress. Klein’s consistent response – despite any difference in the criterion tests from TAAS – remained the dramatic and sudden change in the statistical correlations between what TAAS asserted was dramatic improvement and what Rand researchers flagged as grave question marks.

Finally, in recognition of the uproar his presentation generated, Dr. Klein reduced his ‘cumulative’ response to the concerns thrown his way with this returning to the startling absence of independent verification through statistical analysis of individual student performance:

- “It’s when suddenly it happens (*dramatical closure of achievement gaps – my reference added for context*) and then two weeks later it disappears. That’s the problem. In other words, I would be much more convinced that you accomplished your goals if I came in and gave a test was similar in nature and I got the same result that you did...That’s the piece that’s different here. I tested the same kids two weeks later and it disappeared. Where did it go? It went into think air. Well, maybe it’s because the objectives are different...I doubt it...It’s extremely unlikely. ..It’s extremely unlikely. If you want to bet, I’ll bet you...”

Back in Texas, the reality that Dr. Klein’s presentation was not the result of a full-blown, statistical study gave the political supporters of the Bush Presidential bid and defenders of the “Texas Educational Miracle” the opening they needed to insist – if not demand – the Rand Corporation do that full study. The Rand did the study. The full report with all its charts, tables, and graphs is available.

In effect, the pressure from Texas education and political power brokers and the Rand’s decision to perform the full study took the power of Dr. Klein’s presentation to the June 1999 NAS conference. A complete study such as this is a rigorous process. That study was not completed to October 2000 – literally on the eve of the 2000 Presidential election.

The full study more than validated the concerns addressed by Dr. Klein some 17 months before Presidential election day in November 2000. But here was the dilemma that Texas had presented the Rand Corporation as expressed to me PERSONALLY IN A TELEPHONE CALL WITH DR. KLEIN:

- Release the report in October on the eve of the election and be accused of playing politics.
- Release the report after election day and be accused of letting political cowardice overwhelm institutional integrity.

As always, the Rand Corporation chose to leave no doubts about institutional integrity. The report was released in October and barely made a wave.

The full report used the nationally administered National Assessment of Academic Performance (NAEP) as the academic control to compare performance and gains with the TAAS between 1994 to 1998. Here are a series of bottom-line conclusions and textual remarks from that study:

- **4th Grade NAEP Math:** “Score increases in Texas were almost identical to those nationwide...”
- **4th Grade NAEP Reading:** “The average black student was roughly in the 38th percentile of all Texas test takers whereas the average white student about the 67th percentile. This gap was slightly larger than the difference between these groups in 1994. In other words, the black-white reading gap actually increased during this four-year period. The SAME patterns was present in fourth and eighth-grade math scores.
- **Consequently:** “...In other words, whereas the gap on NAEP was large to begin with and got slightly wider over time, the gap on TAAS started off somewhat smaller than it was on NAEP and then got substantially smaller.

The report wrote “...The large discrepancies between TAAS and NAEP results raise serious questions about the validity of the TAAS scores...” Among the reasons cited:

- “...TAAS questions are released after each administration. Although there is a new version of the exam each year, one version looks a lot like another in terms of the types of questions asked, terminology and graphics used, content areas covered, etc...”
- “...Thus, giving students instruction and practice on how to answers the specific types of questions that appear on the TAAS could very well improve their scores on the exam...”
- “...In short, if TAAS scores were affected by test preparation for the TAAS, then the effects of this preparation did NOT appear to generalize to NAEP exams. This explanation also raises questions about the appropriateness of what is being taught to prepare students to take the TAAS...”
- “...A small but significant percentage of students may have “topped out” on the TAAS. In other words, their TAAS scores may not reflect just how much more proficient they are in reading and math than are other students. If that happened, it would ARTIFICIALLY narrow the gap on the TAAS between whites and students of color... Thus, the reduced gap on the TAAS relative to NAEP may be an artifact of the TAAS being too easy for some students. If so, it would deflate the gains in TAAS scores over time. In short, were it not for any topping-out, the TAAS gains scores in Figures 1 through 3 would be even larger, which in turn would further increase the disparity between TAAS and NAEP results...’

Key conclusions from the report:

- “...**According to NAEP**, Texas fourth graders were slightly more proficient in reading than they were in 1994. However, the country as a whole also improved to about the same degree. Thus, there was nothing remarkable about reading score gains in Texas...”
- **According to NAEP**, “...In contrast the increase in fourth grade math scores in Texas was significantly greater than it was nationwide. However, the small improvements in NAEP eighth grade math scores were consistent with national trends...” (key qualifier follows)
 - “...**In all analysis** including fourth grade math, the gains on the TAAS were several times greater than they were on NAEP.

- “...**The huge disparities** between the stories told by NAEP and TAAS are especially striking in the size of the gap in average scores between whites and students of color and whether these gaps are getting larger or smaller...
 - “According to NAEP, the gap is large and increasing slightly. According to TAAS, the gap is much smaller and decreasing greatly...”

The Rand issued its report. Presidential election day saw the Texas governor take the Texas Educational Miracle to the nation’s capital. No Child Left Behind was on the horizon. Texas’ ultimate judicial victory was months away.

Between Dallas I.S.D. report, Dr. Klein’s presentation at the NAS conference, and the full Rand Corporation report, the fairy tale that Texas was telling was exposed statistically to little effect.

But what those reports had in common were the questions they raised about the academic integrity of the TAAS testing program and answered those questions with solid statistical analysis. What those reports did NOT do was to look at the actual test questions.

Others did and those that did give solid evidence that the statisticians told the truth.

We deliberately skipped to calendar years November 1998 and June 1999 to show the ferocity of the defense that key officials of the TEA advocated supporting and validating the academic integrity of the TAAS testing program. With a January 2000 federal court decision on the near horizon, the TEA had no choice other than this rigorous assessment of academic integrity.

With those defenses, we gave insight into the hard data and the statistical analysis from the second largest school district in Texas and one of the nation’s leading independent research organizations that was mounting to question the integrity of the entire testing program.

Now now step back to 1995 to give a premonition of context to those two settings when a courageous school administrator out of Temple I.S.D. and her team became, in effect, the first whistleblower to directly challenge the TEA’s assertion of TAAS grade level integrity – in this case the entire math testing program.