

- [Brian's Blog](#)
- [Daily Kos](#)
- [DIRELAND](#)
- [DMI Blog](#)
- [Gotham Gazette](#)
- [James Wolcott](#)
- [Majikthise](#)
- [Matthew Yglesias](#)
- [Mother Jone](#)
- [MYDD](#)
- [Steve Clemons](#)
- [Talking Points Memo: by Joshua Micah Marshall](#)
- [The Albany Project](#)
- [The American Prospect](#)
- [The Daou Report](#)
- [The Decembrist](#)
- [The News Blog](#)
- [The Politicker](#)
- [Washington Monthly](#)

• Disclaimer



EdWize is sponsored by the United Federation of Teachers (UFT) as a place where members, public education advocates and others can express opinions in an effort to establish an agora of informed commentary on public education and labor issues. The views expressed here are not necessarily the official views of the UFT, New York State United Teachers or the American Federation of Teachers. Anyone who claims otherwise is violating the spirit and purpose of this blog.

• Spam Blocked

90,273
spam comments

- [New Teacher Diaries](#)
- [Teaching](#)
- [NYC DOE](#)
- [Charter Schools](#)
- [Labor](#)

January 21, 2008

Tweed's "Value Added" Project: Ideology Trumps Education



Filed under: [Teaching](#), [Testing](#), [Education](#) — Leo Casey @ 7:42 pm

New York Measuring Teachers by Test Scores: so reads the headline on the front page of this morning's *New York*

<http://edwize.org/tweeds-value-added-project-ideology-trumps-education>

1/22/2008

Times which announces the NYC Department of Education's secretive pilot project to use value added statistical measures of student standardized test scores to examine the performance of teachers. The teachers and their schools will not be informed that they are the subjects of this study.

The DoE's "value added" project is a fundamentally flawed exercise which can not possibly deliver what it promises. It is being pursued, with the full knowledge of its flaws, because technocratic ideology trumps sound educational practice at Tweed. Moving forward with such a flawed project is extraordinarily irresponsible because "value added" — the idea that one should measure how much academic progress students have made, rather than just their absolute academic standing — holds promise as an useful tool in the repertoire of schools and educators. But the way in which it is being recklessly pursued by Tweed will cast discredit on the entire enterprise.

The DoE has no contractual or legal authority to use test score data in the evaluation of teachers, and the UFT will oppose it with all the means at our disposal. This is a line in the sand for the UFT.

To understand just how intellectually dishonest this exercise is, consider the following. This pilot project is based on student scores on the annual New York State ELA and Math standardized exams, grades 4 through 8. [The initial year of testing, grade 3, provides a baseline, leaving only grades 4 through 8 — the years in which the exam is given on an annual basis — for the measurement of progress.] This means that the pilot can only be applied to the teachers of grades 4 and 5 in an elementary school, and to ELA and Math teachers of grades 6 through 8 in a middle school, a small fraction of all teachers. More importantly, since the ELA and Math exams are given in January, the students will have had at least two different teachers in the interval between exams — one in the spring term of one school year and the other in the fall term of the next school year. Even assuming that the exams are an accurate and complete measure of student learning — and there is ample evidence that they are not — a student's progress from one exam to the next is thus dependent upon at least two teachers. In some instances, a student would have another two teachers for Academic Intervention Services in the 37.5 minute tutoring, and a fifth teacher if he attended a summer program. How could one possibly isolate an individual teacher's contribution to a student's progress using this method?

When confronted with this problem by the UFT and other educators and experts on "value added" it consulted, Tweed decided to move ahead nonetheless, by simply dividing the progress of the student between the two primary teachers. It does not require an advanced degree in statistics but rather simple common sense to understand that this defeats the purpose of the entire exercise. If a student has a really phenomenal, accomplished teacher in the spring term of one year, followed by a struggling novice teacher in the fall term of the next year, her test scores are likely to be flat or even regress, given that the struggling teacher is teaching in the period leading up to the exam. What Tweed's method does is simply divide up the total progress between two teachers who are making very different contributions to the student's academic progress, reverting both to the mean, as the progress attributed to the accomplished teacher is lowered and the progress attributed to the struggling teacher is raised. In short, while Tweed claims that it is measuring the contribution of individual teachers, its method is clearly incapable of distinguishing those contributions.

The defense of this procedure offered by Tweed is that they have done a statistical analysis which concluded that averaging out the progress among teachers provides an accurate measurement of individual teacher contributions. The exact nature of this analysis was never explained; we were supposed to take this conclusion on faith, even as it defied elementary logic. No doubt, if one did aggregate analyses, averaging the contribution would not present a difficulty, precisely because the differences among teachers would cancel each other out in the aggregate. But the claim made on behalf of this project is that it will accurately measure the **individual** teacher contribution, and that is clearly not possible when one can not differentiate between the contributions of two or more teachers to an individual student's progress.

To those who have followed the development of "value added" models of measuring academic progress, this fundamental flaw comes as no surprise. Experts such as Bill Sanders, senior research fellow at the University of North Carolina who devised the first value added model for education in Tennessee, point out that in their current state of development, these models are unrefined tools for individual differentiation and specification, at best identifying the outliers — the very best and very worst performers. And that is when there is one to one

correspondence between a single teacher and the period of a student's preparation for the exam. According to Sanders, value added models provides the most accurate data when one is looking at aggregate categories, such as the growth of teacher skill over time. [He finds that, on average, teacher skill increases through the first decade of performance, and then plateaus.] Sanders has said he won't participate in a value added project which is not done in collaboration between the school district and its teachers, and he has had nothing to do with the New York City pilot.

One serious problem with the use of value added models for individuated, disaggregated analysis is that in order to perform such exercises, one must assume that which is clearly not the case — that students are randomly assigned to different classes and teachers. Princeton economist Jesse Rothstein has done a statistical analysis which shows that this assumption is false, using an elegantly simple falsification hypothesis — a fifth grade teacher should not have any effect on a fourth grade test score. The fact that a statistical analysis shows precisely such a relationship is explicable only by the fact of school life that every teacher understands, that students are not randomly assigned to classes and teachers. This is one very important reason why value added models produce meaningful statistics on an aggregate, rather than an individual, scale: if you consider all of the students in a school, you have eliminated the problem of the uneven distribution of students among classes and teachers, and if you consider all of the students in a district, you have eliminated the problem of the uneven distribution of students among schools.

We are now in the midst of an era of the great misuse of standardized exams. The experts who design such exams — psychometricians and psychologists — are outspoken in their insistence that exams are designed for different purposes, and that it is entirely illegitimate to take a test designed to diagnosis problems in a student's reading comprehension, for example, and use it to reach a judgment on whether the student has mastered the skills she needs in English Language Arts, or to make a high stakes decision about a student's promotion or graduation. With this misuse of standardized exams, the DoE is now attempting to extend this illegitimate use of standardized exams to make high stakes decisions about teachers, a purpose for which they were clearly not designed.

What is remarkable about Tweed seizing on such a fundamentally flawed and intellectually dishonest project is what it says about its estimation of the ability of its administrators, after six years of Klein's Children First agenda. Millions upon millions of public and private dollars have been spent on its much vaunted Leadership Academy, but now Tweed is looking for ways to circumvent the professional educational judgments that the graduates of that academy make on teaching quality. There is a technocratic ideology that guides all of Tweed's ventures, one that saw the development of ARIS — its new multi-million dollar computer database — as a way to directly observe and evaluate everyone in the system from a computer terminal in Tweed, without having to rely upon any intervening human judgment. In the hands of Tweed, ARIS is eerily reminiscent of the panopticon [literally, all seer] of the 19th century English utilitarian Jeremy Bentham, a 21st century 'virtual' version of Bentham's prison architecture which allows an omniscient power to track each and every subject. Tweed's technocratic ideologues are so intent upon having that power in the numbers on their desktop computer terminals that they will pursue that goal even in the face of the knowledge that those numbers can not possibly be accurate and complete.

Teacher evaluation is the subject of collective bargaining and teacher tenure is a matter of state law. The UFT will not open our agreement to consider any role for such a fundamentally flawed project in the evaluation of teachers, and we will defend with all our means the tenure law. Just this last year, the state legislature passed and Governor Spitzer signed a law which laid out three grounds for making decisions on teacher tenure — [1] supervisory evaluation, [2] peer review and [3] the ability of teachers to use data to inform their instruction. That is as it should be. Teaching is a demanding, difficult craft that is not reducible to a technocrat's numbers.

[Permalink](#) [TrackBack](#) [Share on Facebook](#)

5 Comments »

Comments are open for registered users and do not reflect the views of the UFT. Please read our [general rules for commenters](#).

- Leo,

today's Times says:

The United Federation of Teachers, the city's teachers' union, has known about the experiment for months, but has not been told which schools are involved, because the Education Department has promised those principals confidentiality.

Can you clarify something? I am assuming that the UFT knew the DoE was collecting data, not that they were plotting to evaluate teachers.

Jonathan

Comment by jd2718 — January 21, 2008 @ 8:44 pm

- [...] Chester Finn, originally strong supporters of market forces have had second thoughts and Leo Casey at Edwize skewers the Department plan. If we drive "merit" dollars to teachers who the [...]

Pingback by "You Gotta Keep the Devil Way Down in the Hole," Fighting Poverty and Tweed « Ed In The Apple — January 21, 2008 @ 9:54 pm

- The DoE told us that they wanted to do an "academic study" on value added models over the summer. Randi made it clear from the very first time this was raised in a meeting that we would be in total opposition to the use of data from that study to evaluate teachers. When they told us how they wanted to do the study, we raised the sort of objections I discussed above, to no avail. The spectre of using this study for teacher evaluations and tenure was raised by the DoE first last week, in a speech the deputy Chancellor gave in Washington DC.

Comment by Leo Casey — January 21, 2008 @ 10:31 pm

- It says in the times that the UFT knew. What exactly does that mean? The rank and file certainly did not know? So why were we not told? Secrecy is NOT the same as leadership.

Comment by dr_dru — January 21, 2008 @ 10:32 pm

- Dr. Dru:

Perhaps you should read the comment immediately above yours before you write.

Comment by SOC ST TEACHER — January 22, 2008 @ 10:26 am

[RSS feed for comments on this post.](#) [TrackBack URI](#)

Leave a comment

You must be [logged in](#) to post a comment.

• Meta

- [Register](#)
- [Login](#)
- [Entries RSS](#)
- [Comments RSS](#)
- [WordPress.org](#)