

## The Irreconcilables

Matthew Di Carlo

Senior Fellow, Albert Shanker Institute

The New Teacher Project (TNTP) has a new, highly-publicized [report](#) about what it calls “irreplaceables,” a catchy term that is supposed to describe those teachers who are “so successful they are nearly impossible to replace.” The report’s primary conclusion is that these “irreplaceable” teachers often leave the profession voluntarily, and TNTP offers several recommendations for how to improve this.

I’m not going to discuss this report fully. It shines a light on teacher retention, which is a good thing. Its primary purpose is to promulgate the conceptual argument that not all teacher turnover is created equal - i.e., that it depends on whether “good” or “bad” teachers are leaving (see [here](#) for a strong analysis on this topic). The report’s recommendations are standard fare - improve working conditions, tailor pay to “performance” (see [here](#) for a review of evidence on incentives and retention), etc. Many are widely-supported, while others are more controversial. All of them merit discussion.

I just want to make one quick (and, in many respects, semantic) point about the manner in which TNTP identifies high-performing teachers, as I think it illustrates larger issues. In my view, the term “irreplaceable” doesn’t apply, and I think it would have been a better analysis without it.

The report includes performance data for four large districts and one charter management organization (CMO). In the four regular public school districts, “irreplaceable” is defined in terms of growth model estimates (for the CMO teachers, it is based on overall evaluation ratings).

To TNTP’s credit, in these four districts, they do attempt to account for the imprecision in the estimates - e.g., by employing confidence intervals (if only more states and districts [were doing likewise](#)). But, even so, right off the bat, most of their estimates (three out of four districts) are based on only one year of data, and while these scores are [hardly useless](#), it’s not quite appropriate to draw any [strong](#) conclusions about teachers’ effectiveness with such small samples, and that includes grand labels such as “irreplaceable.” For instance, a decent-sized proportion of these teachers will not make the “irreplaceable” cut the following year, due mostly to error rather than “real” change in performance.

To illustrate this (albeit somewhat crudely), I used the New York City “Teacher Data Reports” (described [here](#)) to code teachers as “irreplaceable” according to a rough approximation of one of TNTP’s district-specific definitions (“District B”). Based on single-year estimates in math and reading, a full 43 percent of the NYC teachers classified as “irreplaceable” in 2009 were not classified as such in 2010. (In fairness, the year-to-year stability may be a bit higher using the other district-specific definitions.)

Such instability and misclassification are [inevitable](#) no matter how the term is defined and how much data are available - it’s all a matter of degree - but, in general, one must be cautious when interpreting single-year estimates (see [here](#), [here](#) and [here](#) for related analyses).

Perhaps more importantly, if you look at how they actually sorted teachers into categories, the label “irreplaceable,” at least as I interpret it, seems inappropriate no matter how much data are available.

For example, in “District B,” it is teachers with at least one median growth percentile rank (e.g., in one subject) above 65 and none below 35, while in “District D,” it is teachers with at least one statistically significant, positive percentile

rank and none below the median. In “District C,” teachers are coded as “irreplaceable” if they have at least one score significantly above average, and none significantly below average.

I would characterize these (test-based) definitions as “probably above average” (though remember that teachers need only score highly *in one subject* - half/most of a teacher’s estimates can be statistically average - and those in Districts B and C can actually have most of their point estimates *below* the mean/median - so long as one of them [e.g., one subject] is discernibly above).

In other words, calling them “irreplaceable” seems, at best, an exaggeration.

Now, I fully acknowledge that there is no widely-accepted definition of concepts such as “irreplaceable,” and that such definitions inevitably entail subjective judgment (mine is on display in this post). In addition, as stated above, I give them a lot of credit for paying attention to error in constructing their definitions.\*

But I think it might have been a better analysis had TNTP avoided the overblown, media-friendly characterizations of their performance categories. They could have made their points and presented their results, many of which are interesting and meaningful, without doing so. There’s a big difference between characterizing teachers as “higher-performing than average” and saying they’re “nearly impossible to replace”

Many of their readers might not be aware of the issues involved, and the report itself is very light on discussion of them.

(I would add that a bunch of the core results are based on surveys of teachers. I’m [all for querying teachers’ opinions](#), but, while TNTP does have a large sample, the survey is voluntary. This is not their “fault,” but it does mean that the results cannot necessarily be used to make generalizations about what the teachers in these districts think.\*\*)

Don’t get me wrong -there are certainly “irreplaceable” teachers, and there’s no doubt that they often leave the profession for reasons that can be partially prevented. TNTP has a viewpoint on how to do this, and ***none of the discussion above speaks to whether their recommendations are good or bad.*** But, when it comes to the characteristics, attitudes and behavior among teachers who are “nearly impossible to replace,” I would interpret the results of this particular report with a healthy dose of caution.

\*\*\*\*\*

\* My *guess* is that TNTP predetermined that roughly the “top 20 percent” of teachers should be classified as “irreplaceable” in each of the four districts (perhaps in part to provide a sufficiently large sample to use their survey data), and then calibrated their definitions to produce that result. Yet - and remember I’m speculating here - the scores in some districts were so imprecisely-estimated that they couldn’t achieve the 20 percent sub-sample without relaxing their definitions to the point where they (at least in my view) no longer reflected “irreplaceability.” This may be why, in “District A,” for example, where three years of data were available, the bar is higher (e.g., teachers cannot have any estimates below the mean) than in the other districts - because the more precisely-estimated value-added scores allowed for a more stringent identification of “top performers.” If data availability varies between districts, there is no shame in acknowledging that the ability to use these data to identify high-performing teachers might also vary. Also, seeing as the term is so central to the report, it would have been impressive for them to have presented results for one or two alternative definitions of “irreplaceable,” to see if they were different.

\*\* TNTP does not seem to report response rates, perhaps to protect the confidentiality of their districts, but they do say they required 20-30 percent in a given school, depending on the district. Another thing that would have been helpful is a full set of survey results, disaggregated by district and performance category. For instance, Figure 3 represents compelling evidence that the high-performing teachers differ in attitudes, but in order to evaluate this, one would really need to see the responses for other, similar questions (if there were others).

Original commentary available at: <http://shankerblog.org/?p=6372>