

Richard J. Murnane, Harvard Graduate School of Education

Comments on the “Race to the Top” Rules

I would like to offer two comments on the proposed “Race to the Top” rules and regulations.

1. Learning about the consequences of these remarkable investments

The “Race to the Top” initiative will markedly alter the rules under which states and local school districts receive federal funds to improve public education. It also will dramatically increase the amount of such funds. It is critical to learn what the consequences of these significant rule changes and investments will be. This can only happen if research efforts are planned before the money is distributed AND if a requirement is put in place that states receiving “Race to the Top” funds must use a small percentage of the funds to pay for high quality evaluations of the consequences of their initiatives, and that these evaluations be conducted by an external group of researchers. There is precedent for such a requirement. When the federal government began to allow states to experiment with welfare reform initiatives in the early 1980s, a requirement for receiving a waiver to use federal funds to support these welfare program initiatives was that money be put aside for random assignment evaluations of their consequences. These random assignment evaluations provided a tremendous amount of fine-grained information that was critical to the design of the Welfare Reform Act of 1996. Without the federal requirement for the random assignment evaluations of welfare initiatives during the 1980s, little systematic knowledge would have been available to inform the design of the 1996 welfare reform legislation. The “Race to the Top” initiative provides a similar opportunity to dramatically increase knowledge of the consequences of new educational initiatives. This opportunity will only be realized if the requirements for receipt of “Race to the Top” funds include high quality evaluations that are designed before the money is put to use. IES has the capacity to create review boards to evaluate the quality of the research designs proposed for the evaluations.

2. Using student test scores to evaluate teachers' performances

Several provisions of the proposed "Race to the Top" eligibility rules include the use of student test scores to evaluate teachers' performances. When done well, fitting value-added models that provide such performance estimates will be useful in guiding the decisions of local school districts. Using best-practice methods to fit estimate value-added models using high-quality longitudinal data on individual students does provide valuable information. The results are especially useful in distinguishing between classrooms in which a great deal of learning appears to have taken place during a school year and classrooms in which very little constructive learning seems to have taken place. (The models are much less effective in distinguishing among classrooms in which a moderately amount of student learning seems to have taken place..) However, I suggest two cautions. First, the skills required to specify and estimate value-added models that incorporate "best practice" in this research area are not widely distributed. Nor are the longitudinal data needed to do a good job in fitting such models widely available. The requirement that states make use of teacher performance estimates based on value-added models in order to be eligible for "Race to the Top" funds is likely to lead to the proliferation of estimates that are of much lower quality than can be obtained with best practice methods and high quality longitudinal data. The second caution concerns the sensitivity of results of these models to many factors, including the choice of tests. A high quality research study by Harvard doctoral student, John Papay, entitled "Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates across Outcome Measures" (currently under review at the *American Educational Research Journal*) shows that the rank ordering of urban elementary school teachers based on estimates of their effectiveness reading was remarkably sensitive to the choice of the test used to measure students' reading skills and to the timing of the test. The work of J.R. Lockwood of the Rand Corporation as well as that of other researchers also demonstrates the fragility of the estimates of teacher performance based on value-added models. In my view, well-estimated value-added models based on high quality longitudinal data are valuable. However, the results need to be used cautiously. In particular, the results of the value-added models identify classrooms in which children appear to have made very little academic progress, as measured by test score gains, and classrooms in which children appear to have made remarkable progress. Identifying these classrooms systematically is a significant contribution. However, both sets of finding should elicit efforts to find out why these atypical results occurred. Those doing the follow-up research should

recognize that there are many possible explanations for both apparent exceptionally good performance and apparent exceptionally poor progress. The reasons for the exceptional results should be discovered before policy actions are taken. My guess is that the requirement that recipients of “Race to the Top” funds make use of student test scores in judging teacher performance will elicit a wide range of responses. Some responses will be constructive; others will not. Learning about the factors that distinguish the constructive from the dysfunctional responses is critical to improving the education provided to the nation’s children. This is why it is so important that recipients of “Race to the Top” funds be required to commission the design and implementation of high quality evaluations of the consequences of their initiatives.