A Realistic Approach to Better Testing

Eric Hanushek continues his discussion with Deborah Meier on Bridging Differences today.

Dear Deborah,

Whenever there is any discussion of accountability, there is always an overhang of issues of testing and the use of tests. Testing clearly falls into the knee-jerk reaction zone, but I'm hoping not to inspire such reactions. Instead I want to lay out a simple idea of how to deal with a number of the current problems and objections in order to see if you agree with it.

Let me just give a quick overview of where I think testing under accountability stands. While ultimately we may differ on the weights attached to different observations, that is not central to what I propose. This is just meant to motivate the idea that I want to set out and not to be the subject of any debate.

From the scientific evidence, I conclude that the introduction of testing into accountability has had generally positive effects. In particular, test scores in grades subject to accountability have risen nationally. They have also exposed some embarrassingly large achievement gaps by race, ethnicity, and income levels that have helped to focus attention on the most vulnerable populations. And, as crude as these measures of achievement are, there is now overwhelming evidence that they measure skills that are rewarded in the labor market and that affect the viability of our national economy. In simplest terms, I conclude we are better off having these measures and entering them into the policy discussion than previously when we did not have any measures of student performance.

I think nonetheless we agree that these test scores are narrow and rudimentary, generally failing to measure the higher-order skills that are increasingly seen as important. Moreover, considerable attention has been given to the problems of teaching to the test, to the occurrence of cheating episodes in different cities, to possible distortions in teaching that come from the testing regimes, and to the application mainly at the bottom end. This has led to considerable debate—particularly in the blogosphere—about eliminating testing versus keeping it. I don't think this has been a very productive debate because we are unlikely to end testing and the associated accountability.

And now the point—I don't think that we have to be stuck with the current problems with our testing. I indeed have a very simple proposal to which I would like to know your reaction.

It starts with developing a large item bank of test questions of varying difficulty. Imagine 2,000 questions for 4th grade math that cover the entire scope of appropriate material from basic to advanced topics. Next, make all of the test items—not just sample items—publicly available and encourage teachers to teach to the test, because the items cover the full range of the desired curriculum.

Making the items public will also ensure the quality of the test items. One could invite feedback ratings or open-sourcing to provide a path to improving the questions over time.

Then, move to computerized adaptive testing, where answers to an initial set of questions move the student to easier or more difficult items based on responses. This testing permits accurate assessments at varying levels while lessening test burden from excessive questions that provide little information on individual student performance. Such assessments would not be limited to minimally proficient levels that are the focus of today's tests, and thus they could provide useful information to districts that find current testing too easy. Students would be given a random selection of questions, and the answers would go directly into the computer—bypassing the erasure checks, the comparison of responses with other students, and the like that have followed various cheating episodes.

This proposal actually follows the current testing by the Federal Aviation Administration of knowledge needed to obtain a private pilot license. While there are commercial books on these tests, replete with questions and answers, the efficient way to prepare for the tests is simply to learn the underlying concepts. It is not to attempt to memorize the answers, because it is easy to confuse such an attempt.

What are the potential problems? Some say developing test items would make this too costly, but remember that it is only necessary to have one item bank, not the continually changed banks of today. Some think ensuring that sufficient computers were available in all schools, but with all of the digital devices currently in use, surely there are a range of possibilities to deliver the tests effectively and efficiently. There is the problem that the testing companies would not particularly like the proposal. They find they are happy with mindlessly developing slightly different variants of existing

tests for different states, years, and administrations. But maybe there are more productive ways for them to enter into the process.

The proposed system would yield quick and reliable feedback on student achievement, would deal with the various cheating and gaming issues, and would more effectively define what students should know than the currently available standards.

Rick