

Introductory Remarks by Cassie Guarino

Thank you for coming and participating in this conference. We greatly appreciate your presence, and we very much appreciate the fact that many of you have come from afar.

The goals of this conference are twofold. First, we wish to raise awareness regarding the strengths and weaknesses of different methods of computing teacher performance measures from student test scores. Second, we want create an opportunity to discuss how to disseminate this information in a way that will enable school systems to make informed choices. We believe that the more information that comes to light, the better chance we will have of allowing the best practices to emerge in the policy arena.

I want to place a little context around this conference. Research on teacher value-added has grown greatly in the past few years. In 2008, Doug Harris, Adam Gamoran, and Steve Raudenbush held a very helpful conference on value-added models at the University of Wisconsin at Madison. Jeff Wooldridge was one of the keynote speakers. I attended the conference because I was beginning a research agenda in this field. Jeff, Mark Reckase, and I began our collaboration shortly afterward. Several other people were joining the initial group of researchers in this field around that time, as well. Many of these people are present here today.

Now, five years later, it's a perfect time to take stock of what we have learned. In fact, I think it's essential to do so at this time. As we are all aware, the push for accountability in public schooling has extended to the measurement of teacher performance, accelerated by federal efforts through Race to the Top.

At this point in time, a large number of states and districts across the country are computing measures of teacher performance based on the standardized test scores of their students. Some of the policy makers who are currently on the front lines of implementing these measures in teacher performance evaluations are here today.

The growing field of research on value-added has therefore occurred simultaneously with fast-paced growth in the implementation of these procedures. And, in addition, there has been an increasingly competitive market for assessments and derivative products that compute teacher effectiveness measures.

With the implementation of the Common Core Standards in the majority of states and the development of specific sets of assessments that align with them, it is particularly important that we open up the field right now to talk about the best way to compute teacher performance measures based on the evidence that we've been compiling.

There are many important questions to ask about using student test scores to evaluate teacher performance. A first basic question is whether statistical analyses

of these scores can give us valid and reliable information about teacher effectiveness. No experiments can be designed to test teacher effects, so the answer to that question is virtually unknowable. A few studies, however, such as those conducted by Kane and Staiger and Chetty, Friedman, and Rockoff, have offered encouraging evidence regarding the validity of the measures.

Many of us believe that such measures are valuable but have to be used with caution. There is no research that suggests that we should use these measures to make high stakes decisions with regard to individual teachers without taking other information into account, because even if these measures may be valid, they can still be imprecise.

A second basic question is how best to compute these measures. When a state or district decides to construct teacher performance measures, there are many methodological choices to be made.

At this point in time, research has uncovered a lot of information about the strengths and weaknesses of different approaches. It is important to recognize that not all approaches to computing these measures are of equal quality. Some approaches are stronger than others.

This conference is about discussing these different approaches, so that we can promote the best practices in policy implementation. We will talk about many different methodological choices: the estimators to use, how to test them, and how to implement them. We will also talk frankly about some remaining unknowns in the estimation of teacher effects. While research has provided the answers to many questions, some still remain unanswered.

The goal right now is to do the best we can with what we know.

This conference is organized to maximize interaction across research-policy lines

- Research presentations with Q&A
- Panel discussions
- Break-out sessions
- Time for conversations

We wish to encourage you to participate as actively as you can, to meet as many people as possible, and to engage in conversations that will lead to the sharing of information.

Before we begin our first research presentations, I'd like to take a moment to express my gratitude to everyone who has made this conference possible.

- IES—regrettable that our program officer Phill Gagne is not able to participate due to the government shutdown
- coPIs
- Speakers who have agreed to come and share their knowledge and ideas
- All those of you who have registered and are here in attendance
- Students on our team:
 - Andrew Bibler
 - Eun Hye Ham
 - Michelle Maxfield
 - Brian Stacy
 - Paul Thompson
 - Kelly Vosters
 - Steve Dieterle
 - Francis Smart
- Staci Sharp