

Carnegie Knowledge Network • What We Know Series on
Value-Added Methods and Applications

Webinar 13: How Can Value-Added Measures Be Used for Teacher Improvement?

Q&A with Susanna Loeb

January 15, 2014

Q: Let's start with a question on using value-added for improving hiring decisions. Are there any examples of large urban districts using the value-added of their current teachers to identify effective teacher preparation programs to improve their hiring in the future?

A: That is a good question. Some states are using value-added measures as part of evaluations of their teacher preparation programs. Louisiana has been working on this as the most recent phase of their teacher preparation accountability system. Tennessee has done some work with this as well, creating value-added measures for many of their teacher preparation programs. New York City conducted an analysis that was not specific to preparation programs, but did look at the value-added of different characteristics of teachers that they could judge before hiring. They gathered measures at the point of hiring through different types of screening and saw how those predicted teacher value-added later, trying to validate whether the measures were useful measures for selection.

Q: We have a question on the potential for misuse of value-added. Do you see the potential for wide spread misuse of value-added in policy applications? For instance, what is the potential that a district that fires the lowest 5 percent of teachers fires some teachers wrongly identified as ineffective?

A: Certainly, there is a potential for misuse. Misuse is easy, because the measures are so simple. They are numbers; they feel like something real. To a certain extent they are real; they measure student test score gains, adjusted for background characteristics. On the other hand, there may be better information available. If I had to make a decision - for example, if I had to lay off five percent of teachers because of budget cuts - then I would want to think about how to do this with the least cost to students. Value-added may be useful for that or there may be other measures that are available at the time that are more predictive of how effective those teachers are would be in the future at helping students – and therefore more useful. Most likely, value-added measures would be part of a group of measures that could help me to best predict which teachers are most important to keep. If I had a lot of other good measures, then value-added measures might not be useful. I doubt that there would be a situation in which it would be best to use value-added measures alone without other measures, though there could certainly be a case where value-added measure didn't help at all. Value-added measures may be particularly tempting because they are a measure created from student test scores and that may make decision makers more prone to use them than they should be given their shortcomings.

Q: This question is related to factors outside of the control of the teacher. How does No Child Left Behind's emphasis on tutoring for low performing schools impact value-added results for general classroom teachers?

A: That is, again, a good question. If tutoring is given to students or if an aide comes into the classroom and the aide is very effective, or if, similarly, a principal sees that one teacher is really struggling and puts a coach in with that teacher, then value-added measures may look much better than they should

Carnegie Knowledge Network • What We Know Series on Value-Added Methods and Applications

Webinar 13: How Can Value-Added Measures Be Used for Teacher Improvement?

given the teacher's contribution to student learning. These additional inputs are one reason why evaluation programs should use value-added in combination with other kinds of information and allow for some subjective judgment. Instead of basing decisions on value-added measures alone, it seems to me to make more sense to give principals information such as value-added scores, but then allow the principals to make decisions on their own with knowledge of the context, including what kind of supports were the students receiving from other people aside from the teacher. Considering context is an important part of using value-added well. Maybe eventually the data will include enough information about students' sources of support outside of the classroom that value-added measures could include adjustments for these differences, but we certainly aren't close to the capacity to do that well now.

Q: We have a question about the actionability of some of the findings you suggest here. Is there enough information out there to use value-added in the three ways you described? What more might be needed?

A: It depends on what alternatives are available when you have to make a decision. As pointed out in a different Carnegie Knowledge Network brief, there are a lot of decisions that have to be made. Whether they are made by just ignoring the decisions or whether they are made explicitly, those decisions about whom to hire, whom to promote, what professional development to offer or require are unavoidable. The question is whether there is enough information to know whether value-added measures are useful. That is, does the availability and use of value-added measures lead to better decisions? The existing research provides evidence that value-added measures give some information about teacher effectiveness. Having a teacher that is higher value-added, on average, will benefit students in the long run. There is also information out there that shows that value-added measures are imperfect measures. Teachers with the same value-added, can have very different long-run benefits for students. So, value-added measures are imperfect but there are decisions to make. The question in practice is whether the value-added measures lead to better decisions given the other information that is available.

Q: We have some listeners interested in how we can use value-added to target professional development. Can you explain how value-added could be used to target professional development for teachers? Can you provide guidance to districts that would like to tie value-added to professional development?

A: In the simplest example, consider if you have a teacher whose value-added score indicates that he or she is having trouble helping students to learn the material. That low value-added score could then trigger a more in depth assessment of that teachers' needs which could include specific professional development.. While all teachers may benefit from professional development, we may want to use a disproportional amount of resources on the least effective teachers because of a concern for that teacher's students. Value-added measures can help identify these teachers.

Q: You mentioned that the knowledge base on the statistical properties of value-added is well-established among researchers like yourself, but it is less well understood how it would be used in

Carnegie Knowledge Network • What We Know Series on
Value-Added Methods and Applications

Webinar 13: How Can Value-Added Measures Be Used for Teacher Improvement?

practice. We have a hypothesis here that suggests that the imprecision of value-added makes it difficult to get teachers to believe that value-added is an accurate portrayal of their teaching and that their belief seems necessary for programs of performance incentives to work. What do you make of this?

A: I think it depends on what your performance incentives are based on. My own hypothesis is that it is very hard to incentivize someone to do something when they do not really know what it is they are supposed to do. Because of that, incentives around more concrete measures that teachers can work towards, such as practices outlined by observational protocols, whether perfect or not, are much easier to respond to than are incentives based on student outcomes. Teachers in an evaluation system based on observations can say, "I can do that and someone is going to come back and see if I am doing it, so I am going to try to do that." It is much harder for a teacher to know how to change somebody else. Improving students' test performance is quite an amorphous goal. My sense is that it is more promising to have incentives that are based on concrete actions that people can respond to easily. The research evidence, however, is not clear on this. We have some evidence that incentive systems that were solely based on performance have not been particularly effective in the United States, while some of the more recent ones that involved both continuous feedback and continuous assessment on concrete actions have led to some kind of improvement.

Q: We have some questions about the international studies that you cited. This question asserts that the articles you mentioned from India and Kenya are not straight forward examples of how value-added might be used as an incentive for teachers. Evidence from those studies suggests that teachers did not change their practices, pedagogy, or homework assignments, but instead conducted more test preparation sessions instead. What can we learn from these studies about school improvement where test scores went up but it was not necessarily based on instructional changes, but on test preparation?

A: There is some variation in that as well. It was true in some of the studies that teachers spent more time on test preparation, but in some cases they also had more class time, more homework and other indications of changes in practice. I do think that the conclusion in the United States is not positive, so I am not sure that I am the person to argue for how we can make performance pay based on value-added an effective strategy. I think that if we have it we might want to look at whom it attracts into the profession. I am actually not convinced it would attract people who are more effective into the profession but we really have no evidence on this either way. I do not think performance pay is a particularly promising or transformative approach, though some other people think it is.

Q: We have a question about what we can learn from value-added to apply to all the other teachers who do not have value-added. Since the majority of teachers do not have value-added scores, how can we address hiring and other issues of consequence using value-added and other measures?

A: This question points to the potential of value-added in the first of the three uses in the policy brief - assessing the effects of different programs or approaches on teacher effectiveness and student learning.

Carnegie Knowledge Network • What We Know Series on Value-Added Methods and Applications

Webinar 13: How Can Value-Added Measures Be Used for Teacher Improvement?

We use value-added where we have it to see whether teachers with certain kinds of skills or practices are more effective in terms of contributing to student test score growth. We then would need to extrapolate from this information about the characteristics of effective teachers with value-added scores and whether these characteristics are also likely to signal more effective teachers for those subjects and age groups for which value-added is not available. For some teachers the extrapolation likely makes sense and for some other teachers, it may not. Some of the practices that are included in these observational protocols are really quite similar across different kinds of teaching. This similarity is one indication that some of what it takes to be a good teacher is similar across tested and untested areas. In those cases we can learn from value-added measures, what knowledge and skills are common to more effective teachers. Researchers can also provide evidence on whether the extrapolations make sense. On a small scale, researchers can collect data on student learning linked to teachers for ages and subjects where such measures are not regularly available. They can use these measures to find out about the characteristics of more effective teachers. For example, in kindergarten where you would not want to use typical assessments, you might want measures of child development and see what characteristics of teachers contribute to those measures.

Q: We have a question about the role of the federal government in the use of value-added for improvement. What is the role of the federal government in using value-added in the three ways you presented?

A: The federal government has focused much more on the use of value-added for human resource decisions than on the other uses of value-added that I talked about. This federal emphasis on value-added for human resource decisions is especially evident in places that have received Race to the Top grants. Nonetheless, even with the current focus on human resources decisions, the federal government has been instrumental in facilitating the other uses as well simply due to their role in making student tests score available at such a large scale. These scores have provided lots of opportunities for learning about how to support students and teachers for greater educational opportunities.

Q: Do “best” value-added models exist? What states or districts have the best value-added model in your opinion? Are there some “best practices” to note in design or model choice?

A: Strong controls of student’s prior achievement and, preferably, the prior achievement of other students in the class are the most important features of the good models. If multiple measures of students’ performance prior to entering the class are available and included in the model, those are going to be stronger models. After that there is some evidence that ones that have students’ fixed effects are less stable, so you probably do not want to use those. Among the ones that are commonly being used in practice, I am probably not the one to assess if there is a one “best practice,” but I do not think that there is. All of the ones that are currently used are quite similar. Unfortunately, different approaches will have different implications for individual teachers that are very close to cutoffs; but the value-added scores will be quite highly correlated.

**Carnegie Knowledge Network • What We Know Series on
Value-Added Methods and Applications**

Webinar 13: How Can Value-Added Measures Be Used for Teacher Improvement?

Q: We have a question on the comment you made on the correlation between value-added and years of experience. Does the observation that teachers are more effective in their second year than their first, and more yet in their third than their second, argue against the use value-added and suggest that using years of experience as a proxy for teacher quality would be a good idea?

A: Actually no. The increase in value-added with years of experience does demonstrate that, on average, teachers improve. However, experience differences still explain very little of the differences in value-added across teachers. As an example, if we divided new teachers into five groups based on their value-added in their first two years of teaching, even after five years, the bottom group will not have caught up with the initial performance of the middle group, on average. This is to say that while on average teachers get better over time, the variation within a given level of experience is much greater than the variation across experience levels. In particular, new teachers who begin with high value-added are likely remain high-value-added; not all of them, clearly, but a lot of them.

Q: How unique is the United States in using student achievement growth measures to try to make human resources decisions or improve programs? Do other nations collect value-added type measures and what do they do with them?

A: I do not know what data that is available everywhere, but most countries do not have students take tests every year and do not link those tests to teachers. This approach is rare and is relatively new in the United States as well. So no, it is not common internationally, even among the highest performing nations.

Q: Do you know if other countries are catching on to what we are doing? Have you engaged with some of these other countries and gotten a sense of what their opinion is on how we are doing teacher evaluation?

A: I want to make sure to emphasize that value-added for teacher evaluation is not the only way that value-added can be useful. There are more promising uses. I do think that there is interest internationally in using value-added to understand what good teaching practices are and what good programs are, and potentially to understand about individual teachers. Some countries are collecting data that follows students over time and can link them to teachers and are trying to use the data to create measures of teacher effectiveness. No country that I know of is doing this on the scale that we have in many states. This said, even in the United States value-added is not being used that much for evaluation and when it is being used for evaluation it is in combination with other measures.

Q: What are ways to ensure the evaluation frameworks minimize the risks that you mentioned?

A: We have a brief on this as well. My own sense is that this again is a question about the use of value-added in practice and that we need more information on how to use it well. My guess is that there are promising approaches such as using multiple measures and embedding some subjective judgment in addition to formal measures. A second issue is not about the measures, but about the decisions being made. Given the imperfection of all sources of information, it makes more sense to make many, many

Carnegie Knowledge Network • What We Know Series on Value-Added Methods and Applications

Webinar 13: How Can Value-Added Measures Be Used for Teacher Improvement?

small decisions that each has small consequences for teachers, than it is to have a single high-stakes decision. If you make many decisions along the way and learn about individual teachers from the results, then you are less likely to run into the problems that you will clearly run into if you just measure value-added at one time and base a teacher's whole career on the measure. Single high-stakes decisions are clearly not the optimal way to run a human resources system, to retain highly effective people and encourage them to improve.

Q: We have a question about variation in scores across these different measures. One observation that has gotten attention recently is that value-added scores have a much bigger range and regular distribution than other measures, such as observation protocols. What does this tell us about value-added and those other measures?

A: That is right. Value-added measures tend to have a bigger range than other measures. The observational measures tend to cluster at the very top of the rating scales when used in practice. My co-author and I saw this in study where we asked principals to give ratings that would remain confidential to a set of individual teachers. We then compared those ratings to the ratings the teachers received as part of the evaluation system. The confidential scores had a much greater distribution than the official ratings, which were largely confined to the top two levels. The evaluation ratings still provided some information. That is, the variation between the highest and second highest score was predictive of the ratings on the larger scale. However, the information was clearly muted. I can imagine reasons for the high scores on the official evaluations – whether the principals want to keep the teachers so they don't want to provide negative reviews or they just want to avoid conflict. Value-added measures don't have these issues because they are based on student test-scores. Now some of the variation in value-added measures is just noise, so not all the greater variation is good; but some of it, likely, more fully captures the range of teacher effectiveness than the clustered ratings.

Q: If value-added is to be used for teacher improvement, teachers must understand it. From your experience, do teachers understand it or is it seen as more of a black box? Do you know of any states, districts, or vendors who have done well at helping teachers understand what value-added is?

A: I certainly have talked with teachers who feel that value-added measures are a mystery, a black box. I think that the easiest way to explain value-added measures is that they are the average test score gains of students in the classroom adjusted for the measured background characteristics of students. I have heard anecdotally that it can be easier to explain percentile growth measures than regression-based measures, but they do not seem to me to be easier to explain. In the end, I think it is less about the measure itself and more about how it is explained. I don't know of particular states or vendors who have done a good job of this.