

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

Webinar 10: What Do We Know About the Use of Value-Added Measures for Principal Evaluation?

Q&A with Susanna Loeb

October 9, 2013

Q: Isn't the domain of control problem that principals do not have control over all the factors that lead to student outcomes also relevant for teachers or is it more of a problem for principals?

A: My impression of this is yes to both of those questions. It is also an issue for teachers, but it is an even greater issue for principals. Because teachers actually have the students in their classroom, they directly influence those students. Moreover, we are better able to compare teachers who are serving similar students and have likely similar issues that come up. So by comparing classrooms that are really quite similar in the characteristic of students, we are able to adjust for some of those things that are out of the teacher's control in a way that we have a much more difficult time doing for principals.

Q: What other instruments for principal evaluation have you reviewed and what do you think the validity and reliability of these measures are?

A: That's a very good question. The ones that we have reviewed are the ones that are available on a large scale in the datasets that we use or create. This means that they are not the most subtle measures. They are things like: we surveyed principals and had them assess themselves, we had all assistant principals evaluate their principal, and the district does their own evaluation. None of these are very well tested for validity, in part at least, because we do not really know what they would be tested against. There are certainly measures of principal effectiveness that have been tested more than the measures that I have used. For example, there is the Val-ED measures of principal effectiveness. I am not sure to what extent that has been validated, but that is a more careful measure. The situation for measures of principal effectiveness is even worse than it is for teachers. We have put a lot of research into trying to understand the properties of value-added measure for teachers, but for other measures we have much less strong assessment of their validity and reliability. So your question is a good one. My answer is that we have looked at a range of different measures, none of which have very well-assessed properties.

Q: Would you say the same for surveys of school community, parents and students, and community culture?

A: That's right. The MET study has done a little bit to compare student surveys to teacher value-added measures and other measures of teacher effectiveness, but I don't know of strong tests of the validity or reliability of most of the climate surveys that districts have implemented. Moreover, many of these surveys do not have strong response rates so they may not be representative of their population.

Q: Are there lessons we can draw from other industries or professions on supervisor evaluations? Do other industries use value-added type evaluation in this way?

A: I am not an expert in this. I did some reading in preparing the brief. My sense is that there is a wide variety of different evaluations used. In many industries you are not using value-added measure for the vast majority of the staff because it is hard to tie individuals to outcomes. For managers of organizations, sometimes outcomes such as the revenue or profit of the organization is part of the evaluation and that clearly is only partially in the control of the leader who is being assessed. It is very

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

Webinar 10: What Do We Know About the Use of Value-Added Measures for Principal Evaluation?

clear that it is the performance of the firm and not necessarily the performance of the leader, but they are still evaluated on it. It is a good question of why that is. Even though it is imperfect, it may be useful because we do not have better measures. It may be useful even if it isn't better than other measures because it signals the organizational goals.

Q: When controlling for variables that principals have some degree of control over, you are also collecting a degree of error? How confident are you that we can isolate principal effect from error? Are we just collecting error measures and calling it value-added?

A: Good question. Principals turn over quite a bit. In general, in many large urban areas about 25% of principals turn over each year from a school. When you have turn over so regularly, your measures of improvement come over very short periods of time. My sense of that is that there is too much noise in the measures of principal effectiveness based on school improvement to capture real principal effectiveness; it is just too noisy. Now, if we had multiple outcome measures and could somehow combine them maybe we could get rid of some of the noise, but on just the test performance measure it seems to me that there is too much error there to take it seriously. On the other side, there may not be too much error in the measures of principal effectiveness based on school effectiveness instead of school improvement. We can get a sense of school effectiveness while a principal is there. But with measures of school effectiveness, the potential for bias is probable greater. It isn't clear how much of school effectiveness to attribute to the principal.

Q: We have a suggestion about a fourth approach to principal evaluation. It seems that you could get around the small comparison set problem in approach two by comparing principals that serve similar schools. What are your thoughts on this?

A: The school effectiveness approach is trying to do that: comparing the schools by controlling for things like the characteristics of the students. So it's not just the average test score gain of students but the average test score gain relative to what we would have predicted it would be given who the students are. We can also control for characteristics of the school such as size, grade level, or neighborhood characteristics. So the first approach – principal effectiveness as measured by school effectiveness – is really comparing each school to somewhat similar schools. If there were ways of identifying schools other than things like the characteristics of the students that the institution had, they could be used. The district, for example, could say we think these schools are similar, let's just compare among those schools. That would be a nice alternative.

Q: What does research say about using value-added for principal evaluation in primary schools versus secondary schools? Are there important differences that we should be aware of?

A: I don't think that the research is quite there yet. There is research that shows that the job of the principal is different in those two different kinds of schools. The usefulness of value-added, whether the value-added is more useful in one situation than the other, is more difficult to say.

Q: Different school models envision different roles for a principal and different relationships between principals, staff, and students. For instance, different schools may see the principal as an instructional leader, a CEO, a disciplinarian, community relations person and, therefore, the role of the principal

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

Webinar 10: What Do We Know About the Use of Value-Added Measures for Principal Evaluation?

will differ. How does this affect the validity and reliability of value-added assessment for principal evaluation?

A: My first inclination is that this difference, while interesting to think about, is not a reason that we should have a different approach. To a certain extent it depends on whether what the principal does is the result of a choice that he or she made or whether the principal was forced to act in that way. If the only lever that principals have available is discipline, then we should not hold them responsible for all the other levers that could have improved the schools. But if principals choose this kind of approach versus another kind of approach, then they should be held responsible for these decisions and the same type of evaluation that applies to one is more likely to be useful for others. One principal may choose to spend more time observing in classrooms while another may spend more time developing structures in the school to help to support instruction. It seems to me that we can still compare them even though they made different choices.

Q: Wouldn't some principals have more short-term effects on student outcomes, and others have long-term effects based on the school and management style and what they choose to do? How can value-added, or more generally a principal evaluation system, account for these potentially varying effects?

A: These are really good questions and ones that we still have to struggle with, both for principals and for teachers. My guess is that accounting for both the short- and long-term effects will require using multiple measures. You have the value-added that will have to do with student learning. If you take that simple model, that could be a measure of school effectiveness which we want to attribute a little bit to a principal, but to other things as well. But then some things are much easier to affect in the short-run than student learning. For example, attendance is easier to affect, satisfaction is easier to affect, and retention of able staff is easier to affect in the beginning. It may take a while for other reforms or the other changes the principals make to come through. Having multiple approaches to measure principal effectiveness is then one way of getting at the different ways that principals affect schools. For teachers, one of the great benefits of value-added measures is that they provide a consistent measure across a large number of teachers. It is difficult to collect such a consistent measure from principal evaluations of teachers, for example. However, there aren't nearly as many principals as there are teachers. As a result a district may be in a better position to collect more subtle measures and multiple measures that better capture the context that principals face and the approaches that they take. Even in these situations knowing how effective the school is while the principal is there may be useful for the evaluation. It may be less useful to have some sort of formula that applies equally to everybody.

Q: There seems to be awareness that we should only compare principals working under similar conditions, but don't principals in the same district control the same things, including funding, curriculum, class size, etc.? If so, could we compare principals to their counterparts in same districts?

A: I think yes, that is true. It is easier to compare within a district than between districts. Even within districts principals' jobs could be different if they come into schools that are different, for example, in the skills of their staff. One may get a school with a highly functioning staff; and the other, a school with substantial conflict. How much should you be held responsible for the different performance of the two schools? Certainly if a principal is there for ten years, there is an argument for holding her responsible. But in the first year, the school with the lower functioning faculty, that just might be a more difficult place to get the same kind of school effectiveness, even if you have the similar type of control over

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

Webinar 10: What Do We Know About the Use of Value-Added Measures for Principal Evaluation?

resources. Let's say both principals are able to hire whomever they want, but they cannot get rid of teachers. Then in some ways they have the same control, but in practice they face very different contexts.

Q: I believe the rule of thumb regarding years of data to apply to teacher value-added evaluation is at least two, preferably three. How many years of data would you recommend using for value-added evaluation for principals?

A: At this point, I'm not sure, even with three years of data, the extent to which we should be attributing school effectiveness to principal effectiveness. With teachers, I worry most about imprecision in the measures, so more years of data, particularly a second year of data, is helpful. For principals, I'm more concerned about bias, though imprecision may be important also. Two years of data, of course, is always better than one, and three is better than two. Principals affect more students than teachers do, but our ability to adjust is much lower. So, schools are sometimes actually more unstable than teacher value-added. However, any information on how well a school is doing is useful, even if it is just based on a single year. It just should not be the only source of information that you use to evaluate a principal. It should be considered in context.

Q: You mentioned that much of the principals' affect on test scores may be indirect, so why would we use value-added under this assumption? Should we consider only those direct measures of principal effect? What might be the case for using such an indirect measure of principal effectiveness?

A: For that question, I would actually give a similar approach as with teachers. In what teachers do, I am calling it direct relative to what principals do, because the teachers are interacting directly with the students. But if you look at it conceptually, the teachers are giving instruction and it is what the students are actively taking from it that makes the students learn or not. Learning is an active process, so why not evaluate teachers on what they do, instead of what happens to the kids. I think there is something valid about that: for one, it helps teachers know what to do better. But there are good things about using test performance, because student learning is an outcome that we really care about. We do not really value instruction for itself. The value that we give instruction is about all the good that it does. That is why using measures based on student outcomes is compelling. I think that same notion applies to the principals. It is good to look at intension, culture, actions, and all these other things. But in the end, it is really the outcomes for students that we care about. So having some link to that, both to signal the importance of students and to see the extent to which the principal or teacher might be affecting students is worthwhile to consider as part of an evaluation system.

Q: We have a question about learning objectives, which are becoming more prevalent. Has any research been done around using student learning objectives in principal evaluation?

A: Actually, not that I know of, though I'm sure principals set goals for their schools with district leaders that they then try to work towards. I haven't worked with any of these measures.

Q: This question is about principal effectiveness over time. We use current year test scores to calculate teacher value-added because teachers mostly have students for only one year, but a principal will have an effect over multiple years on the same student, as long as students are in the

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

Webinar 10: What Do We Know About the Use of Value-Added Measures for Principal Evaluation?

same school. Why don't we use an aggregate measure of test scores over that entire time period for that student to evaluate the principal?

A: That's right. If you have a principal who is with a student for multiple years then it is better not to just look at one year. I agree that it would be useful to consider the multiple year effect. If the principal has an equal effect on students in each grade, then measuring the effects as we have been talking about, year-by-year, is fine. If the principal has a different effect in each grade, then we need to think about the approach more carefully. In a paper that Jason Grissom, Demetra Kalogrides, and I wrote in relation to this, we tried a measure of principal effectiveness that controlled for test scores of students prior to entering the school for middle and high school principals. If we had had assessments from prekindergarten, we could have controlled for that for elementary principals. It did not turn out to be that much different in practice in the one example we did, which is why I did not emphasize it here, but it could be a better approach in some situations.

Q: Where do you see this field going? I imagine when you first started doing some teacher value-added research, you may not have anticipated that it would have reached the scale in policy and practice that it has reached today. How quickly is value-added for principal evaluation growing? Where do you see it going?

A: We are actually going to have another brief that talks about this stuff a little bit for both teachers and principals, so this is a nice lead into that. One use of these evaluations is not actually to evaluate and judge individual principals, but to understand which policies or practices are working well and which are not. For example, we can see whether a group of principals who received a given training seem to have greater growth in the school. The problem that we see with the growth measure, for example, is that for an individual principal it has a lot of error, but maybe if we looked at groups of principals that experienced some intervention we could learn about the effectiveness of that intervention using the improvement measure. So one place for value-added is as a tool for understanding programs and approaches that work, rather than as an evaluation. Another place where it may be useful is in combination with other measures. What research needs to figure out is not so much whether value-added measures are inaccurate or biased, but in what cases these measures can be used productively for school improvement. You can have a very accurate measure that is just really bad to use because it leads to lack of collaboration or other behaviors with adverse consequences. Alternatively, there may be situations in which imperfect measures can be used for the benefit of schools and students. So where my world is going is to improve our understanding of how to use the information we have available or can collect in ways that will improve schooling.