

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

Webinar 14: Is Value-Added Accurate for Teachers of Students with Disabilities?

Q&A with Dan McCaffrey and Heather Buzick

February 7, 2014

Q: Do you have any data on how student achievement growth is impacted by having a different accommodation from one year to the next?

Heather: I have some state level data that distinguishes between different accommodations. We are seeing that there are a couple types of accommodations that are more associated with jumps in growth or decreases in growth. The read aloud accommodation in particular is one type of accommodation that a lot of research has shown leads to increases in test scores. Extra time is another accommodation with positive results. If, for example, a student had extra time last year and maybe the read out loud this year, that would not be something that we could disentangle to see because the changes could be from either one of those accommodations. We could look at each of those independently, so we could look at just students who had read out loud this year but not last year or student who just had extra time this year but not last year to get closer to disentangling those results.

Q: To what extent do district cutbacks in special education provisions affect value-added models for teachers of these students? Students may have the same nominal accommodation from year to year, but cutbacks may impact the implementation of it.

Dan: I think that this is related to the previous question. On paper they get this accommodation, but we do not necessarily know if they actually receive it. I do not have any specific data on this. Heather, is your data detailed enough?

Heather: No, but I am hoping that the new data might be. In some cases, there will be electronic databases as the tests become computerized where they can actually tell if students are receiving their accommodations. That will be very helpful.

Dan: The other issue is that students with disabilities are allowed the accommodation, but the student may not use them. The data reporting is inconsistent. Sometimes the data is reporting the accommodation received, sometimes the data is reporting the accommodation that is available. Those do not necessarily equate. That detailed data would be useful, but we are still gathering it.

Heather: One other useful piece of information would be item-level data of accommodation use. Every accommodation would not be as effective for all item types. If we could have detailed information about the specific accommodation used by item, we could do more research to think about how accommodations impact the measurement of student growth.

Q: What do we know about the adverse incentives of controlling for student disability status? Is there good reason to believe that value-added is causing special education referrals for low performing students? This already may be happening, but is value-added increasing this problem?

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

Webinar 14: Is Value-Added Accurate for Teachers of Students with Disabilities?

Dan: It is something that people are concerned about. From a theoretical perspective on what we know about how people respond to incentives and people's understanding of how special education might work, there is an incentive there. But I have not seen any data where someone has looked at referral numbers before the implementation of a value-added state-wide evaluation system and looked at them following the implementation, especially one that is well known to control for disability status within the model. In some ways the wide spread use of value-added as part of formal evaluation is relatively new, so it may be a little early. I do not think that anyone has looked at it, but it would be valuable to research. We do not have any data, that I know of, to show that it is occurring, but we have lots of reasons to think it could occur.

Q: We have a modeling question about controlling for inclusion versus self-contained classrooms. To control for disability, would a useful proxy be to control for student assignment to an inclusion class as opposed to a self-contained class?

Heather: That could definitely be a possibility, but as far as the data that we have, we do not have information about that currently. We have had to estimate if students are in a self-contained classroom or if students are in a general education classroom.

Dan: If we are able to get information at that level, we may be able to get information about a student's level of disability and much more detailed information.

Heather: Some of the data we are working with is several years old. And we know what teacher a student has been assigned to just by a piece of paper stapled to the test. We definitely need better data.

Q: We have a question about teachers' opinions on some of these issues. What do teachers with significant proportion of students with disabilities think of these issues? Do you have a sense of their general attitudes for controlling disability status or including these students in value-added?

Heather: I have spoken about this topic with several audiences and have not received too much feedback. They have been very interested in it, but I have not heard anything negative about it. A few years ago I spoke with a parent of a child with Downs Syndrome and she thought it was important for her child to have rigorous standards for his development and growth. I would consider that one of the issues. As far as controlling for the model, I have not heard too much. Value-added has been in the newspapers a lot of the past couple of years. There have been some quotes about English learners and students with disabilities in classrooms and teachers feeling like it is not fair to receive low value-added scores because of this, but I have not seen anything directly about controlling the model for students with disabilities.

Q: We have a question about Student Learning Objectives (SLOs) as a different way of attempting to assess student achievement growth. Could you speak to the development of SLOs for teachers of students with disabilities, especially in regard to fairness for teachers? What types of SLOs might be appropriate?

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

Webinar 14: Is Value-Added Accurate for Teachers of Students with Disabilities?

Dan: I do not have a lot of expertise on SLOs. I think there would be an expectation that because the SLO is developed with the specific knowledge of the kind of students in a teacher's class that it could be customized more accurately to come up with objectives that are attuned to the specific needs of students. So in a classroom with a wide range of student needs, such as a special education classroom, it allows faculty to be more attuned to that. Exactly what those things would be, I do not have that level of detail or expertise. The problem with those types of measures is how comparable they would be from one classroom to another classroom and whether meeting those objectives in a special education classroom would be the equivalent to meeting it in a regular ed classroom. I think that this is one of the real challenges of SLOs in general. The SLOs may remove some of the concerns that we have about value-added because they could be attuned to matched the students in both a regular and special education classrooms. On the other side, they can be so non-standardized and so different that it will be hard to verify or validate if they are comparable measures. I do not have any specific details, but I suspect that it is something that many people will be considering as part of the process for these teachers, in particular because many special education students are going to be part of the evaluation system for teachers. One of the big challenges of SLOs is balancing customization and comparability. As a system we will have to keep thinking about how to meet that balance.

Q: In the presentation you talked about multiple teachers who may be working with students, particularly special education students. Are there any evaluation systems that take into account the different roles and responsibilities of different staff who contribute to student learning, not just teachers, but also counselors, speech and language therapists, etc.?

Dan: Not that I am familiar with. In the systems I am familiar with when they consider linking professionals with students in the modeling are linking teachers.

Heather: There are some systems that give a school-wide value-added score to teachers in non-tested classes. This may include some of the staff that you have asked about. There are obviously issues with this since it is highly weighted on the teachers who are actually teaching tested subjects, like math and English, but there are some states that are doing that.

Q: How many standardized tests scores do you need to measure a low-performing student's performance with reliability?

Dan: I have not done that calculation specifically, so it would be hard to say. I do not want to venture a guess as it would depend on the test. Some state's tests and certain subject tests are more reliable on the lower end of achievement than others. It may also be that, depending on how low a student is scoring, the number of tests would be really quite high. If students are really taking tests were they are giving very little information then you would need quite a few of them. I do not really have a precise number.

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

Webinar 14: Is Value-Added Accurate for Teachers of Students with Disabilities?

Q: When you speak about the different types of tests that may impact the reliability differently, do we have some reason to be optimistic that computer adaptive tests will increase the reliability for students at the lower end of the distribution?

Dan: Yes, I think we do. That is one of the real promises and motivations behind computer adaptive tests. The idea with computer adaptive tests is that the test a student takes is more attuned to where their achievement level is. So a student who is lower-achieving gets more questions that are challenging for them, but not impossible. The same case goes for students at the higher end; they receive more questions that are challenging for them, rather than too simple. This way, students at the lower end are receiving more reliable test scores. In cases where I have seen the reliability of computer adaptive tests relative to other tests, you get much more comparable reliability across the whole range of students. You do get students at the very ends of the distribution whose test scores are somewhat less reliable. Yet, it really evens out the reliability in comparison to the standardized, fifty-question tests that students are taking.

Q: We have a question about using valued-added for making consequential personal decision for special education teachers. In your opinion are there any issues or concerns we should have for using value-added as an indicator for making hiring and firing of special education teachers? Are there any legal implications under IDEA or any other legislation?

Dan: I cannot comment on the legal issues. In terms of the issues we need to worry about. I think that there are concerns that we should have. As I mentioned value-added for special education teachers may not be comparable. If the models we are using are not particularly good models then we could bump down a lot of special education teachers into lower performance level. It is not a guarantee and a lot of states that are building value-added models are building pretty good models and trying to avoid these biases. I think that it is really important to look at special education teachers across the whole state as a group and get a sense of where they are fitting in the distribution to see if as a class they sit lower in the value-added distribution. I think that value-added can be a useful part of the evaluation as it can for any teacher. It has important information about how their students are achieving. But whenever we are using value-added we need to think about the quality of the information we have from the value-added in terms of making an inference about the teacher. When looking at special education teachers, I think that the quality of the information is less than it is for other teachers. They typically have fewer students because they often teach fewer students and more of their students do not have test scores for every year, so they wind up having a smaller sample of students. This results in more bobble from year to year so pooling multiple years of data is going to be more important for special education teachers. There is a potential that if the model is off that we are undervaluing those teachers. Maybe we should be looking at their value-added, but be careful about applying it in the same way that we use it for all other teachers. It may work or may not, but we should be monitoring and making sure we are being safe. The tests may not measure their contributions as well as they do for other teachers. We should think about developing alternatives for students' achievements. We could develop a study to see if you test special education students with a different type of test would a teacher's value-added look more promising.

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

Webinar 14: Is Value-Added Accurate for Teachers of Students with Disabilities?

Thinking about ways to do some of those kinds of studies would be useful. Value-added can provide information, but we need to be more cautious with using this information for special education teachers.

Heather: I think that Dan had some really good points. I just want to add that for general education teachers as well, value-added should not be the only measure that we use to make such high stakes decisions. There should be other measures and thought that goes into these high stakes decisions. We do not have value-added scores for all teachers, so there is an issue of fairness in that. Then there are all the other complications that Dan has raised today. The monitoring of value-added and using multiple measures is important, especially in high stakes decisions.