

## Interview Transcript: Evaluate outcomes to show your program is making a difference

- **Geoffrey Borman**, evaluator for the St. Paul Public Schools' 2007 Magnet Schools Assistance Program (MSAP) rigorous evaluation, and a professor of Educational Leadership and Policy Analysis, Educational Policy Studies, and Educational Psychology at the University of Wisconsin-Madison
- November 2008

### Interview Highlights

- There are many factors to consider in deciding whether to conduct an experimental or quasi-experimental evaluation of your magnet program.
- Random assignment is useful for eliminating selection bias, but there are certain conditions you must meet to effectively conduct an experimental evaluation of your program.
- For quasi-experimental designs, it is important to have good student-level data and sufficient numbers of participants.

My name is Geoffrey Borman, and I am the external evaluator on the St. Paul Public School magnet district evaluation. I am from the University of Wisconsin at Madison. I think it's really important to be able to understand what the effect of the magnet program is; that we should use practices in schools, in classrooms, that have a proven track record of success. An experimental design isn't always feasible and isn't always appropriate. First of all, the main criterion is, are you answering a cause-and-effect kind of question? Do you want to know, does implementation of a new magnet program impact student achievement? So, your question has to be a cause-and-effect question. That's the first criterion. The second criterion, I think, is that there has to be an over-subscription or there has to be some form of being able to admit some students to a distinct treatment and some students to a control or traditional program intact.

The problem that a random assignment study really addresses is this problem of selection bias, or selectivity. And it becomes a problem when we have a situation like a magnet school, where students and family self-select into that school. And they choose to come to that school for a variety of reasons—that, perhaps, they are very proficient already in science; they are doing very well in science; or that they are really engaged and really interested in getting the best education possible for their kids. And so, in those circumstances, it's really difficult, again, to know whether or not it's the magnet school that's producing the stronger outcomes or if it's just in something that's inherent to the children and to the families that they bring with them.

And, by virtue of random assignment, we have these two groups that are equivalent on everything that we can measure and on even things that we might not even think about or might not have good measures for, like their motivation, or how much their family is really devoted to science, and how often they talk about it over dinner at the dinner table. We don't have any clear measures of those, but the beauty of random assignment is that, on average, by randomly selecting kids to go to the treatment or the BioSMART program and the control program, they are going to wind up being equivalent when we compare them at baseline. And then, we know that if they are equivalent at baseline—at the entry point—that anything that happens to them after that, any kind of differences that we observe after that, we can much more clearly attribute to the difference in the program, because that will be the only characteristic that will really be different between the two groups.

In Washington, we had the fortune of having these two competing kinds of programs operating within that school, and equal interest in both of them, such that we were able to determine randomly which students went to the traditional program and which students went to the BioSMART program. Oftentimes, it's not feasible to do an experiment. In Arlington, it was a school that really has a low enrollment, and it was a situation that they could fill every seat in the program—that they didn't have what's called an over-subscription, or more students applying to the school than we had seats available within the school. So, we were stuck with a situation where we couldn't pull off that kind of experimental design within Arlington and had to go to a different type of design, which is a quasi-experimental design, where we were comparing all of the children who were enrolled and admitted to the Arlington program to a group of students not attending the BioSMART program, but attending other high schools within the district, who grew up in the same neighborhood, who attended the same middle schools, and who looked very similar to the kids who were attending the Arlington BioSMART program.

With a quasi-experimental design, it's much more important to have very good data coming into the study. And I think just having a wealth of information about what the students are coming into the program with; what skills they are coming into the program with in terms of their past achievement outcomes; what the general characteristics of those students are in terms of their family background and so forth; those are going to be characteristics that, if we have really good measures of them, we can come up with very good matches for those students from other schools throughout the district that are comparable; that we can compare the students in the more traditional schools throughout the district to those who are attending the magnet school.

One of the important things to consider beforehand is the size of the sample that you have; to determine what's the minimum sample size that I need in order to carry out a credible rigorous evaluation that's going to be informative and that's going to have the kind of statistical power that I need to really uncover an effect of a magnet school if that effect is there.

A rigorous evaluation is not an easy thing to carry out, and I think oftentimes it really helps to get outside advice and to seek expert opinion about how to implement a high-quality rigorous evaluation. And I think to seek out advice from external partners who are able to help with some of the design issues in terms of the experiment or the quasi-experiment that you are trying to conduct, but still within the district, bring that substantive expertise about the magnet program and about what it is you really want to find out; that that combination of external advice and expertise, and the substantive expertise within the district, can often create a winning, high-quality, rigorous evaluation.