

Beware! Teacher effects could mess up your individually randomized trial! Or such is the message of this paper focusing on what happens if you have individual randomization, but teachers are not randomly assigned to experimental groups.

The key idea is that if your experimental groups are systematically different in teacher quality, you will be estimating a combined impact of getting a good/bad teacher *on top of* the impact of your intervention.

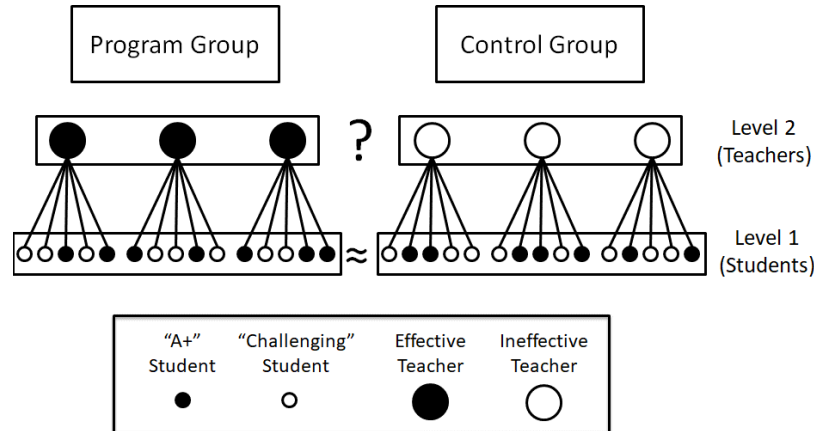


Figure 1: A hypothetical randomized experiment.

Consider the hypothetical experiment illustrated in Figure 1. A mix of “challenging” and “A+” students are first randomized to a program and control group. Randomization ensures the two groups are not systematically different. In a second stage, students are assigned to teachers, three of whom are delivering the program. In this illustration, the program teachers are more effective. The result? Empirical findings will be comparing the program as delivered by effective teachers to the control

condition as delivered by ineffective teachers. Group differences in outcomes (e.g., credits) will be a combination of program effects and teacher effects, and the two cannot be disentangled.

Could this matter even if teachers are not deliberately sorted like in the hypothetical example? Yes. Using information on how much teachers’ influence student outcomes from prior research, Weiss shows that if teachers are not randomized, it can be difficult to establish whether an intervention “works” or whether results were driven by differences in teacher effectiveness.

The paper also includes an empirical example, illustrating the challenge. A large field experiment reported positive effects for college learning communities. The experiment intended to measure the effect of the learning communities *independent of teacher effectiveness*. Weiss’s results indicate that if the learning communities’ instructors were more effective than average (owing to how they were selected/recruited), the learning communities might not have been the reason for the positive effect after all. If the instructors were less effective than average, positive effects might be far larger than originally reported.

This paper clarifies that when designing an experiment we must carefully consider what is the “cause” that’s effect we’re aiming to estimate. If the cause is an intervention as delivered by a special group of teachers, then individual random assignment may be appropriate. If the aim is to understand the effect of an intervention independent of teacher effectiveness (as is typically the case), then we must randomized teachers (or schools). This lesson applies beyond teachers to various service providers, such as advisors, coaches, psychotherapists, etc.