Threats to Internal and External Validity in Health Professions Education Research

Dario M. Torre, MD, MPH, PhD, associate professor of medicine, and Katherine Picho, PhD, assistant professor, Department of Medicine, Uniformed Services University of the Health Sciences

- Internal validity refers to the degree to which inferences can be made about the causal relationship between two variables.
- External validity pertains to whether study outcomes can be generalized across different persons, treatments, outcomes, and settings.
- Internal and external threats to validity can occur during any stage of the research process.
- Below, we describe some of the most common threats to internal and external validity.

Example: Consider a pre/posttest control group research design to evaluate an Internet-based, multimedia cardiac auscultation teaching program (intervention) compared with a traditional lecture-based program (control). The intervention is aimed at enhancing cardiac auscultation skills of medical students.

Threat	Definition	Example
Internal Validity		
History	Events occurring concurrently with treatment may cause the observed effect.	Students participate in a series of grand rounds on cardiac auscultation while taking the multimedia cardiac auscultation program.
Maturation	Naturally occurring changes over time may be confused with a treatment effect.	Students' psychological development and confidence in the use of technology naturally grow during the time of the intervention and positively affect the cardiac auscultation skills of the intervention group.
Regression to the mean	When units are selected for their extreme scores, subsequent scores are likely to be less extreme.	Students with extremely high (or low) scores on a previous cardiac auscultation test are selected for the study. Their subsequent scores after the intervention are likely to be much lower (or much higher for low scorers) than on the previous test; thus, scores regress toward the mean.
Attrition	Loss of respondents to treatment or measurement can produce artifactual effects if loss is systematically correlated with conditions.	Fifty percent of students assigned to the Internet-based multimedia program could not access the Web site and did not complete the program.
Testing	Exposure to a test can affect scores on subsequent exposures to that test, and this change in test scores can be confused for a treatment effect.	By taking the pretest, students are sensitized to a number of murmurs that are later included in the posttest. Their performance in the posttest may be affected by the content of the pretest.
Instrumentation	The nature of a measure may change over time/conditions in ways that may be confused for treatment effects.	A change in the pretest is made during the experiment by including more difficult and complex murmurs and by decreasing the time given to complete the test. The instrument change, rather than the intervention, may affect students' performance.
External Validity		
Multiple treatment interference	The results of one kind of treatment may not hold when that treatment is combined with other treatments.	Students are exposed to real patients with cardiac murmurs in a clinical rotation, and subsequently to a multimedia cardiac auscultation program. The carryover effects between the two interventions may be difficult to separate.
Reactive effects of experimental arrangement	A result that occurs in one kind of setting may not hold in other settings.	The cardiac auscultation proficiency of medical students in a simulated setting may not translate into the same level of proficiency in a real patient setting.
Interaction of selection bias and experimental treatment	The effect of selection bias interacting with the experimental treatment has an effect on the outcome of the intervention.	The multimedia program is given to students who have poor cardiac auscultation skills as opposed to students who have excellent skills (e.g., a group of residents in training vs. a group of experienced cardiologists). The results of the intervention may not be generalizable; they may be biased, according to the different level of skills of the group selected to receive the intervention.