Despite the enormous variety of grant projects, some general principles of social-scientific measurement apply broadly, whether your goal is to assess a curricular, technological, instructional, or programmatic innovation.

First of all, recognize that the process of measurement is more than bean counting. The mere act of planning measurement frequently challenges one to confront fundamental questions about purpose and process (e.g., backward design). And the mere act of measurement often enhances that which is measured (e.g., the testing effect). In other words, recognize that measurement is a productive, integrative part of what you’re doing, not an imposition, distraction, or addendum.

Measurement begins with a hypothesis about the impact of your efforts. In abstract terms, what key effects do you intend? Think broadly by considering impact on content knowledge as well as skills, attitudes, affect, metacognition, interests, decision-making, and so on. For a course, you might intend to increase students’ knowledge of key terms and concepts, cultivate their interest in your discipline, and create an experience they reflect on positively.

After you define impact abstractly, the next step is to define impact concretely. How will you actually measure the hypothesized effects? To operationalize a measure of conceptual knowledge, for example, you might administer a multiple-choice test used by others in your field (i.e., a concept inventory), and use the number of questions answered correctly as your outcome measure.

A good measure is one that is reliable (produces similar results under similar conditions), valid (measures what you think it’s measuring), and practical to implement. Common threats to reliability include not collecting enough data and subjective measurement. To guard against the former, measure early and often; to guard against the latter, consider using (as appropriate) objective measures, multiple judges/graders, and rubrics.

Common threats to a measure’s validity include the limits of introspective access, a litany of human biases (e.g., overconfidence, social desirability, actor-observer, confirmation), and the difficulty of generalization. To maximize validity, avoid asking people to explain their behavior or make inferences about latent mental constructs; and ask instead about their (recent, discrete) behavior, conscious experiences, attitudes, and interests. In course assessment, for example, ask about their satisfaction with the course but not about how well they learned concepts; to measure conceptual understanding, use a concept test. As much as possible, try to align the measurement situation with whatever real-world situation you hope to generalize to. If measuring conceptual knowledge, ask about a novel, real-world situation that would be informed by such knowledge.

Finally, to collect data using whatever instrument you’ve selected (i.e., research design), keep in mind a few key points. First, there are many ways that those collecting data can unintentionally skew results and ways in which human respondents might merely conform to expectations. Second, when possible, collecting data “before” and “after” is almost always much better than only collecting data “after”. Third, always consider some sort of control group or condition that allows you to avoid confounds such as selection bias. Finally, if you’re administering a survey, at least skim these best practices in survey design.

It would be my pleasure to meet with any of you individually to discuss assessment. Stay tuned for announcements about HILT assessment office/happy hours…

Samuel T Moulton PhD / Director of Educational Research and Assessment / Harvard Initiative for Learning and Teaching