EDITOR'S NOTE

Identifying An Education Research Study's Limitations

Timothy F. Slater, University of Wyoming, USA

ABSTRACT

Discipline-based science education research studies face many limitations. One is that study-participants are human beings and apt to be inconsistent in how they respond to educational interventions—even high quality and highly effective ones. The second is that researchers themselves are human and well poised to use study designs and data analysis approaches that yield the most desired results. In the end, simply having "too small a sample size" is a short-sighted limitation. It is author's intellectually pursuing the full range of possible limitations of a study that new insights and new experimental designs can be intellectually created. The discussion of limitations should bring forth ideas and next steps pathways for researchers to follow, making articles more of a conversation and intellectual stimulation of a research trajectory rather than an abrupt ending to a study.

Keywords: Discipline-Based Astronomy Education Research; Discipline-Based Geoscience Education Research; Science Education Research Methods

The oft quoted Nobel laureate physicist Richard Feynman is well known for advocating that the first principle of scientific research is, "you must not fool yourself — and you are the easiest person to fool" (Feynman & Leighton, 1992, p. 343). What Feynman means here is that practicing scientists are often conducting experiments and analyzing data in pursuit of a personal hunch—a highly educated bet if you will—about uncovering a suspected unknown mechanism that explains some observed phenomena. As a result, even the most respected scientists are prone to occasionally misinterpret their data analysis or unintentionally construct an experimental design that yields them the answer they suspected all along, rather than uncovering true factual insight. Internally, deep down, we want our favored explanatory theories, often generated with considerable mental and fiscal efforts, to turn out to be correct. This desire to be correct injects an unfortunate bias into our science which results in errors. In other words, although we try to do science in a systematic, objective, and unbiased way, as it turns out, science is a human enterprise and, being a human enterprise, is subject to the limitations of being, well, human.

It isn't surprising that discipline-based science education researchers are perhaps some of the most prone to "fooling ourselves" researchers on the planet. Many science education researchers have been, or currently are, first and foremost teachers. The most dedicated and enthusiastic teachers have increasing student achievement and attitudes as their primary, overarching goal. Many of the best teachers have tacit, gut-level intuition about how students best learn, and years of experience confirming that they are correct in their assumptions. As an unfortunate result, when the best classroom teachers start to delve into systematic, objective, science education research about how students learn, science education researchers who are former teachers are highly prone to experimentally verifying that they were correct all along when subjecting one of their own ideas about how students best learn.

This effect of needing to prove oneself right about their teaching is amplified by a researcher's own history of teaching. If one has spent a decade teaching hundreds, if not thousands, of students using a particular strategy or paradigm that turns out under the scrutiny of education research to be wrong or ineffective, then the full realization of the vast number of students taught incorrectly can be devastating—so traumatic that one just doesn't want to see it. As a natural result, science education researchers studying how their own students learn too often construct their experimental studies in such a way to vindicate how they themselves have been teaching all along. In other words, we who have experience teaching are naturally inclined to design studies that inadvertently confirm our suspicions. As Feynman says, "you must not fool yourself — and you are the easiest person to fool" (Feynman & Leighton, 1992, p. 343).

Every paper and research report benefit from authors who enthusiastically embrace the required section on study limitations. It is in pursuing the possible limitations of a study that new insights and new experimental designs can be intellectually created. The discussion of limitations should bring forth ideas and next steps pathways for researchers to follow.

Unfortunately, in far too many papers, the limitations discussions are, well, too limited. The most often cited study limitation is small sample size, with the implied suggestion that a larger sample would give the more accurate result. There are at least three problems with citing small sample size as a limitation. The first problem is that every scientist wishes they had a larger sample size, perhaps sometimes even a sample size that encompasses the entire population studied (in which case, of course, it wouldn't be a sample at all). The second problem of sample size is that simply having a larger sample size becomes a statistical problem because at extremely large sample sizes, every statistical difference become statistically significant. In times like these, it is difficult to tell which statistically significant differences are important effects, and which are results of mathematically having too many subjects in the sample. The third problem of sample size is that readers naturally assume that large sample sizes are always better than small sample sizes—yet, in the field of astronomy, for example, the vast majority of traditional scientific papers published are written using a sample size of just a single subject: one star, one Sun, one asteroid, one crater, one variable star—just one (Slater, Slater, Heyer, & Bailey, 2015, p. 67). The bottom line is to shy away on relying on the notion of limited sample size when discussing an experimental design, because it seen as shallow and lacking insight.

A confounding limitation common to most science education research studies is the demographics and source of the study-participants themselves. Far too often, researchers study their own students. This makes sense, these students are captive subjects, convenient to access, and well known to the researcher. Unfortunately, one's easily accessible students rarely represent the larger population of learners, and many paper reviewers require considerable justification about why one should trust the results of studying one's own students. The rule to take home here is that it is always best to study someone else's students, rather than your own whenever possible.

Moreover, traditionally, many science education researchers would disaggregate their data based on easily observable demographics—most notably apparent sex and skin color. As it turns out, these are highly volatile demographics to lean upon. The most egregious being relying heavily on one's skin color as the critical demographic as we now know that socio-economic status is far more important than one's race in understanding the backgrounds, motivations, and experiences that bring study-subjects to their current status (Thomas & Stockton, 2003). As a result, a careful and extended discussion of precisely who the study subjects are is worth doing, so that reviewers and readers can themselves skeptically look for unintended biases in research designs, thereby adding or subtracting weight to the study's implications.

Perhaps the biggest limitation is that discipline-based science education researchers are studying human beings. On one hand, when physicists collaborating with geologists toss a rock through the window to see what happens when it hits the ground, the rock itself is dispassionate about the experience and should, within reason, give the same results each time the experiment is conducted. Those researchers studying human beings, on the other hand, have a very different experience. To carry the analogy way to far, the truth is that if researchers throw a human being to the ground through a window, one has no idea what will happen after that human being gets up and dusts themselves off. Moreover, if these hapless researchers were to do it a second time, the human subject themselves might not have a consistent reaction. In making my point, I'm not trying to incite violence or break any human subjects research rules here, what I'm trying to humorously say is that human beings react to experimentation in often unexpected and inconsistent ways. This variability in human response to stimulus—even when the stimulus is a positive learning event—needs to be accounted for when writing about education research whenever possible.

Like most areas of science, discipline-based science education research is at its heart a human enterprise. That means that researchers themselves are subject to the frailties and faults of the human condition, the study subjects have the right to be inconsistent in how they react to experimental treatments, and that the implications of any study results need to be interpreted with special care. What I am advocating here is that authors writing scientific papers, taking the extra time needed to improve and exhaust all possibilities of one's study limitations is worth taking the time to allocate.

REFERENCES

Feynman, R. P., & Leighton, R. (1992). "Surely you're joking, Mr. Feynman!": Adventures of a curious character. Random House.

Slater, S. J., Slater, T. F. Heyer, I., & Bailey, J. M. (2015). Discipline-Based Education Research: A Guide for Scientists, 2nd Edition. Hilo, HI: Pono Publishing.

Thomas, J., & Stockton, C. (2003). Socioeconomic status, race, gender, & retention: Impact on student achievement. Essays in education, 7(1), 4.

