



[Five Answers About EdTech Experiments: A Response to Benjamin Herold](#)

By [Justin Reich](#) on April 19, 2018 9:11 AM

Benjamin Herold, an *Education Week* reporter, recently published a story about [an experiment conducted by Pearson](#) to determine whether growth mindset interventions improve learning in a software product to teach computer science. They tested these methods using a randomized controlled trial, where some 9,000 students received the messages and others did not. Publishers and researchers have been conducting this kind of growth-mindset research for many years, for instance in 2013 Khan Academy did a similar experiment [involving over 250,000 students](#). In both the Pearson and Khan Academy cases, schools and students using the product were not alerted to this experiment. Several commentators strenuously objected to educational publishers conducting this kind of research ([example](#)), and Herold posed five questions online about the study. I do research on similar kinds of interventions (see [here](#) and [here](#)), so I thought I'd answer his questions, while I work on a my own response to the study. The questions below are all from Herold.

First question: When an [#edtech](#) company adds an evidence-based intervention into its software (here, [#growthmindset](#) messaging such as "No one is born a great programmer. Success takes hours and hours of practice"), is that product improvement, or clinical research?

Every classroom teacher, educational publisher, and instructional designer implements variation in teaching practice over time to improve teaching. There are virtually no actors in education who do the exact same thing year after year, decade after decade. They introduce this variation, examine whether the variation leads to better outcomes, and make adjustments accordingly. We would consider any teacher, publisher, or instructional designer who wasn't trying to improve their practice to be negligent. So everyone is doing "product improvement." In my view, when educators do this kind of

Source: http://blogs.edweek.org/edweek/edtechresearcher/2018/04/ben_harold_an_ed_week.html

improvement, they should do it in such a way that we can learn from it, improve practice, and share what we learn with others.

Every educational software company and publisher will be modifying their products over time to try to improve them; and I'd like to incentivize them to do so in a way that the public benefits from those companies sharing what they learn.

Second, and related: If an [#edtech](#) company is trying such an intervention, is it better (more effective, more ethical) to run it as a research-style experiment, an A/B-style test, or to roll it out to everyone at the same time?

This question has the wrong frame of reference—it suggests that Pearson had a choice between some people getting an intervention versus all people getting an intervention. For every existing educational product, the only choice is to have some people get a new intervention. In 2016-2017, students used the Pearson product and got the old version of it, and the use of that product led to certain educational outcomes. In 2017-2018, Pearson conducted an experiment, where some 2017-2018 students got an intervention and others did not. If they had rolled out the intervention to all 2017-2018 students, there would still be a group of 2016-2017 students who didn't get the intervention.

With any change in instructional design, some students receive the (dis)advantage of the intervention and others don't. This is an inevitable feature of improving instruction.

When companies or educators introduce change within a cohort through a randomized control trial, we have an excellent chance of understanding whether or not the intervention improved learning or not. When we introduce change between cohorts, it's much harder to understand whether variation led to learning improvements.

Change instructional practices is inevitable, and having some students be (dis)advantaged by those changes is inevitable. Right now, the vast majority of changes in instructional practice are done in such a way that we don't know whether or not the changes were a good idea.

Third, what kind of consent can/should be obtained before running such an [#edtech](#) test? What kind of transparency should be expected/required?

We compel students to attend educational programs, and within those educational programs there are educators and publishers who are constantly testing new approaches--with every new teacher, new course, new textbook revision, new software update. Students are constantly compelled to be subjected to intentional variation in instruction.

I'm very concerned about circumstances where we make it harder to conduct good experiments--to implement intentional variation so that the field can learn from changes to instruction--than it is to run haphazard experiments.

Historically, the United States Department of Health and Human Services, which oversees human subjects research, has been concerned about these kinds of problematic incentives as well. Health and Human Services regulations [45 CFR 46](#) govern human subjects research, and they have been adopted by most federal agencies that sponsor research. These are the rules that gave rise to Institutional Review Boards, human subjects research training, informed consent, and so forth. There are several exemptions to the rules governing human subjects research, and [Exemption #1](#) is for research governing regular classroom practices, such as "research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods." At many universities, if I proposed conducting the study Pearson ran, I would be able to convince the IRB that the research appropriately fell under Exemption #1 and I would not be obligated to get informed consent.

Many of these kinds of rules have not been updated for the digital age of [surveillance capitalism](#), and they should be. But if changes to these rules meant that educational publishers could easily do experiments where 2016 students get something different from 2017 students, but they couldn't randomly assign interventions within 2017 students, then I think we'd be harming our ability to conduct research without having a fairer world.

Fourth, does using [#growthmindset](#) messages in an effort to change students' attitudes ("I can get better at this!") and behaviors ("I'll try to solve this problem again!") count as a psychological experiment? Where is the line?

Psychology is the study of the human mind, and learning is a crucial component of the human mind. So every effort to change students' attitudes and behaviors is a psychological experiment. Moreover, all educational environments implicate emotion, affect, and attitudes. Nearly every software product that we use in education indicates when students are right and wrong--through green checks, red exes, words like "Right," "Wrong," "Correct," "Try Again," "Do You Need A Hint," visual markers like stars and points, audio indicators like buzzes and bells. The introduction of each of these elements affects student attitudes, emotions, and behaviors much like growth mindset messages do. One difference with growth mindset messages, is that they have been [studied under circumstances where we have an increasingly clear idea about their short and long term effects on students](#), unlike nearly every other feature in the user interface of learning software.

There is no useful line between learning research and psychology research, every change to a learning environment is a psychology experiment. We should absolutely examine whether or not new interventions may harm students. In the case of growth mindset, [well-designed large scale studies](#) indicate that they provide slight benefits on average, with greater benefits going to more disadvantaged students. However, we know that these interventions vary in effectiveness across different environments, so it's quite appropriate to test them in new contexts, like computer science courses in higher education, before deploying widely in those contexts.

Compelling small groups of students to participate in research with the potential to benefit all students has a long tradition in American education. Since 1969, we've required a small fraction of student every year to take tests as part of the National Assessment of Educational Progress, and what we have learned from NAEP has been crucial in shaping education policy and raising issues like educational inequality.

And fifth, and last for now--can [#edtech](#) be an effective and appropriate medium for promoting social-emotional-psychological changes/improvements in students? Should it be?

This is another question that strikes me as having the wrong frame of reference. Every instructional design element--every textbook page, every worksheet, every teacher

powerpoint deck, every hanging wall poster, every note home to parents, and every feedback message in a education software product--has social, emotional, and psychological implications on students. Every learning experience makes social, emotional, and psychological changes in students--those dimensions are inseparable, and that's what learning is.

These questions should not be framed as "should materials from educational publishers try to affect emotion and attitude or not?" because every educational material--created as OER like on Khan Academy or by a for-profit publisher like Pearson or by an individual teacher for her classroom--has effects on the emotions, attitudes, psychological states, learning, and social relationships of students.

Nor should we ask "Should companies do experiments on students or not?" because there is no version of our world where for-profit companies don't experiment on students--every typo that a publisher corrects in a textbook is an experiment on students. The question is "what kinds of experiments do we want to encourage?" I believe we should encourage experimentation that we can learn from over haphazard experimentation; for instance, we should encourage randomized controlled trials over cohort based trials (experiment within 2017 rather than between 2016 and 2017), and we should encourage teachers, schools, non-profits, universities, and for-profits to share what they learn when they try to improve their instructional materials.

I'll be the first in line to kick for-profit publishers out of K-12 and replace them with publically funded Open Educational Resources, but until that day comes, I hope that for-profit publishers continue to try to improve their products, and I hope that they when they conduct these educational experiments (and every change is an experiment) they choose methods that let them determine whether the change improved learning, and I hope they share their insights with the field.

It's also very clear to me that everyone working on these kinds of experiments in applying insights from social psychology and behavioral economics to education has an incredible amount of public communication work to do. We have not effectively communicated with the public why this research has the potential to be valuable, and **[we have not sufficiently listened to public concerns about our interventions and](#)**

[approaches](#). I might very well have important misconceptions in my answers above, and I'm eager to hear other perspectives. We need more public forums to discuss these questions, and I thank Benjamin Herold for provoking the conversation.

For regular updates, follow me on Twitter at [@bjfr](#) and for my publications, C.V., and online portfolio, visit [EdTechResearcher](#).