

THE CHRONICLE OF HIGHER EDUCATION

Teaching

June 15, 2015

Everyone Complains About Evaluations. A Nobel Laureate Offers an Alternative.

By Meg Bernhard

The list of complaints about how colleges conduct course evaluations is long and seems to keep getting longer. A [survey released last week](#) of thousands of professors by the American Association of University Professors found that student evaluations are losing much of the value they once had. [Earlier research](#) already showed that student evaluations failed to adequately describe teaching quality, and often reflected judgments about an instructor's appearance. But if not student evaluations, what should colleges use to judge the effectiveness of teaching?

[Carl E. Wieman](#), a Nobel Prize-winning physicist and professor at Stanford University's Graduate School of Education, says he may have found an answer. [In a paper](#) published recently in *Change* magazine, Mr. Wieman suggests another form of evaluation: judging professors based on an inventory of their teaching practices. The ultimate measure of teaching quality, he argues, is the extent to which professors use practices associated with better student outcomes.

"It may seem surprising to evaluate the quality of teaching by looking only at the practices used by an instructor," Mr. Wieman wrote in the paper. But he said research over the past few decades had established a correlation between the teaching methods used and the amount of student learning.

Under Mr. Wieman's proposed method, instructors would be required to fill out course inventories using a template developed by him and his wife, Sarah Gilbert, a visiting scholar at Stanford. An instructor would be asked to quantify, for example, the average number of times per course that students worked in small group

discussions.

A number of colleges and groups across the country have been testing Mr. Wieman's teaching inventory. Some observers laud the evaluation system because it requires professors to reflect on and track progress in their courses. But others caution that the system may be too standardized and that it may be difficult to translate the teaching inventory into other fields of study, like the arts and humanities.

'Active' Learning

Nearly 10 years ago, Mr. Wieman began researching and experimenting with teaching methods in science departments at the University of Colorado at Boulder and the University of British Columbia. He also studied work on best teaching practices in math and science courses to determine which categories to include in his inventory. The research literature, Mr. Wieman said, suggests the most effective teaching method for science, technology, engineering, and mathematics courses — the STEM fields — is "active" learning, in which students engage in problem-solving activities during class time.

He said he was confident that his inventory method could be translated directly to engineering, and "modestly confident" that it could be used in fields outside of STEM.

But some experts and observers wonder how portable the approach can be. After all, teaching a course in chemistry is quite different from teaching one on English literature.

For Catharine H. Beyer, a research scientist in the University of Washington's educational-assessment office, disciplines are unique and require different methods of teaching. In an art class, she said, student learning may best be assessed by a more "hands off" teaching approach, in which a teacher assigns a project and then gives students feedback weeks later. "We need to think a lot about what we're trying to teach in each of these disciplines," Ms. Beyer said.

In his paper, Mr. Wieman acknowledged that he had tested the inventory in math and sciences only, and that some courses, like

seminars, labs, and project-based classes, probably would not fit within the rubric. For other fields, the inventory might require some adjustment. "In developing an inventory like this, it's got to really capture all the standard teaching methods that people in that discipline use," he said.

Other observers worry that standardizing teaching practices might detract from an instructor's autonomy. Different professors have distinctive teaching styles, argued Geoffrey D. Sanborn, chair of the English department at Amherst College. "Education is fundamentally about the relationship between the teacher and the student," he said, adding that in each course, faculty members bring unique teaching styles that they've been developing over their careers. Mr. Sanborn said he worried about potentially harmful effects if certain teaching practices became "enshrined" in a department or a university.

Michael P. Chaney, an associate professor of counseling at Oakland University, in Michigan, said that while he welcomed Mr. Wieman's work, certain elements of instruction would be difficult to measure with an inventory.

"There's an aspect of teaching that cannot be taught, and that's the nontechnical skills like personality," Mr. Chaney said.

'Tyranny of Student Evaluations'

Mr. Wieman hopes that adoption of his inventory will lead colleges to put far less weight on the end-of-semester ritual of course evaluations, when students get to anonymously judge their professors on a standardized form. In his paper, Mr. Wieman wrote that one of his goals was to free professors from "the capricious, frustrating, and sometimes quite mean-spirited tyranny of student evaluations."

But Larry A. Braskamp, who has researched faculty assessment and who next month will become interim president of Elmhurst College, said evaluations are "incomplete" without student input.

Ms. Beyer agreed, saying that "it is wrong to not give students a voice in their own learning."

In an email interview, Mr. Wieman countered that he would keep student evaluations if he adopted the inventory system for a course but that he would use them "more appropriately."

But even if Mr. Wieman's system was used alongside other means of evaluation, another concern remains. Mr. Wieman acknowledged in his paper that his system measures "the use of a particular practice, not how well those practices are being used." He again emphasized the "strong correlation" between learning outcomes in STEM courses and teaching methods used, independent of other characteristics of teaching, like active learning.

Ms. Beyer agreed that while active learning in the classroom can work well, that does not give the entire picture. "Just using a teaching methodology alone doesn't ensure that students are learning," Ms. Beyer said. Others wondered whether professors would complete a lengthy survey on their course and whether they would accurately remember specific details about it.

Mr. Wieman insisted that the survey would take little time to fill out.

Diane Ebert-May, a professor of plant biology at Michigan State University, said that professors were likely to exaggerate their classroom experiences. She co-wrote a 2011 paper, ["What We Say Is Not What We Do,"](#) which confirmed that self-reported data about faculty teaching is different than observational data. While not lying outright, professors tend to misremember or inflate details, she said. That issue can be resolved by balancing self-reported data with plenty of observational data, she added.

No matter what the method, Ms. Ebert-May said that more research was needed on how best to gauge learning in college.

"These tools are fine, we can use them, but I think we need to take the step and make sure students are actually learning," she said.

"There is a lot of work to do."

18 Comments

The Chronicle of Higher Education

Login ▾

Recommend

Share

Sort by Oldest ▾



Join the discussion...

wclibrary • 3 hours ago

"For other fields, the inventory might require some adjustment."

Well, yeah.

Thanks for sharing.

1 ^ | v • Reply • Share ›

mcdaniel_community • 3 hours ago

Bottom Line: There is a lot of work to do...implies what we have learned as practitioners of teaching has been woefully inadequate given the decades of C & I and methodology research to which we have been exposed. It seems to me that the highest priority in higher education preparation should be focused on how to effectively teach. Has anyone looked into active learning employed at the West Point Academy...might be a real lesson there.

2 ^ | v • Reply • Share ›

vlghess → mcdaniel_community • 2 hours ago

Though it is nice to be able to order students to study tables and declare them AWOL if they don't show! :)

^ | v • Reply • Share ›

wggray • 3 hours ago

While the idea of replacing or supplementing student evaluations is quite welcome, the proposed solution has alarming implications for faculty autonomy. It would evaluate teaching according to a predetermined template of administrator-approved practices that are "known" to be effective. "Known" to whom? Education researchers? Should disciplines across the university orient their teaching methods around the recommendations of a field that is notoriously fad-driven and has difficulty replicating research results? (See Makel and Plucker, 2014)

Disciplines -- and individual academic departments -- should indeed be thinking seriously about effective teaching. I'm uncomfortable with the idea that a standardized list of "best practices" can be defined at the university level. There's a lot of hoopla about "active learning" right now and perhaps that may produce solid learning outcomes in the natural sciences; but it seems to be accompanied by a huge dose of disdain for other types of teaching that might be very effective in, say, the humanities. Where I am seeing innovative pedagogy percolate through my field, it's happening organically as younger Ph. D.s leave grad school with fresh ideas and methods. Efforts by campus administrators to stamp certain approaches with an official seal of approval can only lead to faculty resentment and an uninspired, box-ticking approach to writing syllabi.

3 ^ | v • Reply • Share ›

wlgoffe → wggray • 2 hours ago

You said, "'Known' to whom? Education researchers?' If you look at the links in the article, you'll see it is STEM researchers, like Wieman. I'd certainly take a look at Wieman's article in "Change:" <http://www.changemag.org/Archi...> as a number of important criticisms are made of teaching evaluations based on extensive research. I'd also look at the inventory itself at <http://www.lifescied.org/conte...>

1 ^ | v • Reply • Share ›

22078549 • 2 hours ago

From the article:

"The ultimate measure of teaching quality, he argues, is the extent to which professors use practices associated with better student outcomes..." also

"...he said research over the past few decades had established a correlation between the teaching methods used and the amount of student learning..."

Sounds to me like the ultimate measure might better be the amount of student learning. In the sciences, this is not impossible to do, but the evaluation overlords tend to avoid such metrics

3 ^ | v • Reply • Share ›

wlgoffe → 22078549 • 2 hours ago

I agree that it would be best to measure learning (and I suspect that Wieman would agree), but as you suggest, it is pretty hard. He points out that in some disciplines such measures can be rare. In principle, he argues that it might well be easier to look at inputs (what faculty and which is correlated in thoroughly studied disciplines in STEM) that STEM researchers have found correlated with learning.

It is well worth reading the Change article -- referenced above:
<http://www.changemag.org/Archi...>

^ | v • Reply • Share ›

kpd1divine48 • 2 hours ago

Having spent nearly 30 years in R&E primarily in K-12 arena, this is of much interest to me. I often teach adjunct courses for colleges and have wondered how this issue is viewed by higher ed colleagues. From my observation, it appears to be a matter of autonomy (or the giving up of same). In other words, in K-12 there seems to be more acceptance of the fact that school systems, or hiring organizations, have some degree of responsibility (prerogative) to define what effective teaching looks like, and therefore, hold teachers accountable for using certain practices. Not sure I have seen acceptance of that premise present in higher ed. Without it, evaluation systems are difficult to get off the ground.

^ | v • Reply • Share ›

gbreton • 2 hours ago

A couple of points:

1. Student evaluations, in my experience, are truly not particularly useful in gauging how much students have learned. However, they do serve as a good indicator of the classroom learning environment, the engagement of the professor in the course, and help identify faculty members (hopefully untenured) that have no place being in a classroom.

2. Perhaps one of the reasons why students seem to do better (and I think marginally at that) in classrooms that implement "active learning strategies" is that the faculty member running those classes are already strongly engaged with their students and working to the best of their ability to get students to learn. Student learning is most likely enhanced by strong interactions between professor and student regardless of teaching method.

3. Poorly implementing any teaching method (even so-called "proven ones") will lead to poor outcomes. I can still manage to make lousy cookies using the recipe on the back of the chocolate chip bag.

3 ^ | v • Reply • Share ›

fisher56 • 2 hours ago

It is great that academics - especially our colleagues in social sciences to whom we probably owe the idea - finally realize that asking students to rate your class is not really a measure of how well you did your job.

From the article - the proposed innovation:

"The ultimate measure of teaching quality, he argues, is the extent to which professors use practices associated with better student outcomes."

Why not look at the outcomes directly? Presumably there is a way to "measure"

...my notion of the outcome directly. I recognize there is a way to measure those, right? Just give department wide tests or simply look at the performance of students (grades, choice of major) after the course being assessed.

Anyway, it may be too late to fix anything. I recently looked at marketing materials for teaching evaluation software. It was full of manager-speak such as:

"..connect your data in 3rd party business intelligence systems.."

"..a variety of assessments for your organization allowing it to better harness information for strategic purposes.."

I have a distinct feeling the target audience were not peer evaluators in the Promotion and Tenure Committee.

Course evaluations are a management tool, period.

^ | v • Reply • Share ›

vlghess → fisher56 • 2 hours ago

As you imply, assessment of learning (properly done and correlated with other measures) would seem to be the best indicator of the quality of teaching. There remain two problems with that approach. One is political: the only way outcomes assessment has made any inroads at all against faculty resistance is by carefully separating it from the evaluation of the faculty.

The other is real (and is why I can't totally object to the "political:"

As we've learned in the K-12 arena, where "value-added" evaluation of teachers fails to take into account the multitude of variables outside of teacher control. Translated to the HE arena, these include: hours available for out-of-class effort (job and family responsibilities, athletics, etc.), the nature of the course, the "gating" of the course (since, in some disciplines, prior preparation is essential to success in the course), etc. etc.

So--good instructional inputs tell us something about the effort and engagement of the instructor (what we tend to find in teaching reflections as part of the tenure process), they don't tell us everything!

^ | v • Reply • Share ›

tuxthepenguin → vlghess • an hour ago

""value-added" evaluation of teachers fails to take into account the multitude of variables outside of teacher control"

You are correct. I can teach two sections of the same class, with both having the same lectures, assignments, and office hours, yet the outcomes are very different. A couple of the differences are the hours they work outside of class and parental status. I don't know how you'd ever control for that.

^ | v • Reply • Share ›

boiler • an hour ago

Faculty members aren't stupid. If you set up a rubric for evaluations, then tell everyone that their scores will be based on their self-reported adherence to the rubric, they'll all make sure that they can report full adherence. Most faculty are creative enough to find things in their courses that they can describe as small group discussions or active problem solving. The result? Lots of perfect scores, and no way to distinguish effective from ineffective teachers.

2 ^ | v • Reply • Share ›

Russian Dictator • an hour ago

"The research literature, Mr. Wieman said, suggests the most effective teaching method for science, technology, engineering, and mathematics courses — the STEM fields — is "active" learning, in which students engage in problem-solving activities during class time."

In Russia, where STEM education has traditionally been at a high level, this active learning has been achieved by assigning students an extensive and meaningful

homework. Class time wasn't used for tutoring. My understanding of the US educational system 50 years ago is that students were likewise treated seriously back then.

15 years ago Mr. Wieman's research in cold atoms was listened to by the whole community of physicists. Too bad he decided to become irrelevant by devoting his efforts to charlatany.

1 ^ | v • Reply • Share ›

Joseph Merola • an hour ago

I agree that there are real problems with student evaluations. I am not sure that they are any worse than they have been because they have never been a good measure of evaluating teaching effectiveness. The big problem that I have with Wieman's suggestion is that this assumes that ALL teachers can be effective with the tools and methodologies used in the studies. In fact, might not the effect of these studies be themselves stilted by the enthusiasm that the proponents of the study have for those 'experimental' teaching methods? I submit that giving those same tools to a wide range of teachers, some will do well and some will not. There is such a thing as needing to match the tools to the practitioner as well as the subject.

^ | v • Reply • Share ›

dlpike • 35 minutes ago

Part of the challenge is differentiating student satisfaction from student learning and from teaching practice. We often conflate them in end of term evaluations and then are, not surprisingly, unhappy with the data. And with all due admiration for Dr. Weiman's late career interest in teaching, I am still troubled by the attention to his work as if no one else had been doing this or advocating for active learning for decades earlier. His discovery was his own.

^ | v • Reply • Share ›

May • 22 minutes ago

Professor Wieman has been working on teaching approaches for over two decades. He is one of the minority that have truly earned the title Professor - most view teaching as very secondary to their research, shall I name a few? He is innovative and he cares very deeply about students, learning, and general wellbeing. It shows.

^ | v • Reply • Share ›

Professor in Ohio • 18 minutes ago

One of my research areas is Student Evaluation of Teaching. The research community has already established that SET is a bias laden method of measuring something other than quality teaching. Certainly portfolios and providing professional development for the scholarship of teaching and learning is an important step in the right direction. There are still ways to provide student input about the quality of instruction without providing surveys that simply are not useful to faculty seeking to improve teaching and learning in their classrooms or to improving student success. Multiple methods for determining quality of teaching are certainly appropriate and portfolios / best practices should be part of that. **@gbreton** - the description of following the recipe on the back of the box not always leading to good cookies is spot on. Providing a teaching "check list" does not ensure teaching and learning quality.

^ | v • Reply • Share ›

