

Do Tests Help Students Learn or Was a New Study Misreported? (##)

January 28, 2011

Do Tests Really Help Students Learn or Was a New Study Misreported?

By Alfie Kohn

The relationship between educational policies and educational research is both fascinating and disturbing. Sometimes policy makers, including those who piously invoke the idea of “data-driven” practice, pursue initiatives that they favor regardless of the fact that there is no empirical support for them (e.g., high-stakes testing) or even when the research suggests the policy in question is counterproductive (e.g., forcing struggling students to repeat a grade).

Sometimes insufficient attention is paid to the limits of what a study has actually found, such as when a certain practice is said to have been proved “effective,” even though that turns out to mean only that it’s associated with higher scores on bad tests.

Sometimes research is cited in ways that are disingenuous because anyone who takes the time to track down those studies often finds that they actually offer little or no support for

the claims in question. (Elsewhere, I've offered examples of this phenomenon in the context of assertions about the supposed benefits of homework – along with details about some of the other ways in which research is under-, over-, or misused.)

Then there's the question of what happens when the press gets involved. It's no secret that the reporting of research is often, shall we say, disappointing: A single experiment's results may be overstated or a broad conclusion may be vaguely attributed to what "studies show," despite the fact that multiple qualifications are warranted. Possible explanations aren't hard to adduce: tight deadlines, lack of expertise, or a reporter's hunger for more column-inches or prominent placement (hint: "The results are mixed at best" isn't a sentence that advances journalistic careers).

Whether ideology may also play a role – a tendency to play up certain results more than others – is hard to prove. But last week I found myself wondering whether the New York Times would have prominently featured a study, had there been one, showing that taking tests is basically a waste of time for students. After all, the Times, like just about every other mainstream media outlet, has been celebrating test-based "school reform" for some time now, and, in its news coverage of education, routinely refers to "achievement," teacher "effectiveness," exemplary school "performance," and positive "results," when all that's really meant is higher scores on standardized tests. The media have a lot invested in the idea that testing students is useful and meaningful.

So we probably shouldn't have been surprised to discover that last week the Times ran a lengthy (30-something-inch) story on the second page of its national news section under the headline "Take a Test to Really Learn, Research Suggests." And it should be equally unsurprising that the study on which the story was based didn't really support that conclusion at all.

(I'm picking on the New York Times because of its prominence, but many other news organizations also featured this article and described the study in similar terms. Other headlines included: "Taking a Test Helps Learning More Than Studying, Report Shows," "Learning Science Better the Old-Fashioned Way," and "Beyond Rote Learning.")

We should begin by noticing that the study itself, which was published online in the January 20 issue of Science, had nothing to do with – and therefore offered not the slightest support for – standardized tests. Moreover, its subjects were undergraduates, so there's no way of knowing whether any of its findings would apply to students in K-12 schools.

The real problem with the news coverage, though, is twofold: On closer inspection there are issues with how both the independent variable ("Take a Test") and the dependent variable ("Really Learn") are described.

What interested the two Purdue University researchers, Jeffrey D. Karpicke and Jannell R. Blunt, was the idea that trying to remember something that has been taught can aid learning at least as much as the earlier process of encoding or storing that information. Their study consisted of two experiments in which college students either practiced retrieving information they'd learned or engaged in other forms of studying. The former proved more effective.

The type of retrieval practice used in the study was an exercise in which students recalled "as much of the information as they could on a free recall test." But the idea of retrieval practice didn't need to involve testing at all. "The NY Times article emphasized 'testing,' which is unfortunate, because that's really irrelevant to our central point," Karpicke told me in an email message. "Students could engage in active retrieval of knowledge in a whole variety of ways that aren't 'testing,' per se." For example (as he explained in a subsequent message), they might put the book

aside to see how much of it they can recall, try to answer questions about it, or just talk about the topic with someone.

In other words, the experiments didn't show – and never attempted to show – that taking a test works better than studying. They were really comparing one form of studying to another.

Then there's the question of outcome. When I said a moment ago that the study showed retrieval practice was more "effective," the most appropriate response would have been to ask what that word meant in this particular context: more effective at what?

In the first experiment, students were asked both verbatim questions and inference questions that drew on concepts from the text they had been given. In the second experiment, they either took a short-answer test of the material or were asked to create concept maps of that material from memory.

The researchers seemed impressed that practice retrieving facts worked better than making concept maps (with the text in front of them) at preparing students for a closed-book test even when the test itself involved making concept maps. But the students were tested mostly on their ability to recall the material, so it may not be surprising that recall practice proved more useful.

I would argue that this result says less about how impressive the method was than about how unimpressive the goal was. Karpicke and Blunt weren't investigating whether students could construct meaning, apply or generalize concepts to new domains, solve ill-defined problems, draw novel connections or distinctions, or do anything else that could be called creative or higher-order thinking. Now if testing – or any other form of retrieval practice – were shown to enhance those capabilities, that would certainly deserve

prominent media attention. But this study showed nothing of the sort. Indeed, I know of no reason to believe that tests have any useful role to play in the promotion of truly meaningful learning.

The main contribution of the articles that were published about this study is to remind us of the importance of reading the actual studies being described. To understand why the description of this one was misleading, try to imagine a newspaper running a more accurate account – one with a headline such as “Practice Recalling Facts Helps Students Recall Facts.”

Back to Blogs