

Enhancing Educational Evaluation by Incorporating Side Effects

FRANCIS SCHRAG

University of Wisconsin, Madison

COVID-19 reveals the way single causes have multiple effects on children. Proponents of evidence-based education, like their colleagues in medicine, need to attend to the inevitable side effects produced by programs and policies. I explain why continuing motivation to learn is one outcome that no responsible evaluation should omit.

One can look at the COVID-19 pandemic as an enormous experiment in deschooling. The results are not pretty. Loneliness and isolation – the result of stay-at-home and social distancing orders – have led to increases in depression and anxiety. During the pandemic, 25% of high school students reported worsened emotional and cognitive health, and more than 20% of parents of children aged 5 to 12 reported that their children experienced worsened emotional health (Panchal et al., 2021, p. 2). The academic results are equally disheartening, especially those pertaining to the least advantaged. According to a Rand panel of educators surveyed in the spring of 2021, whereas in a normal pre-pandemic year, about 9% of Black and Hispanic children would be more than one grade behind in English language arts and mathematics, almost double this, 17%, were behind at the time of the survey (Kaufman et al., 2021, p. 34). Some might see this as a reason to eliminate assessments during a pandemic, but, of course, it is those very assessments that permit us to take stock of the damage. Whether these dismal results should be communicated to students as failing grades on a report card is another question, one to which I will return below.

These results reveal something else, something we all know but can easily forget: assessments of student accomplishments capture not only students' own capabilities and the effort they put into their schoolwork but the background conditions in their homes and communities that either propel them forward or impede their academic progress. For example, parental job insecurity, lack of a quiet space to concentrate, or limited access to the Internet will likely increase stress levels within the family, reducing or even eliminating students' ability to concentrate on their schoolwork. So, many students are penalized for factors far beyond their control, while others are the beneficiaries of conditions they had no part in producing. Here again, while this can provide fodder for arguments against grading individual students, it is difficult to see what might replace it. So long as our society requires an increasing proportion of adults with highly trained capabilities, selection and sorting for advanced training must be done on some basis or other, and those opposed to using school grades or national tests as such a basis must provide a fairer alternative. Do we want to depend on educators' subjective impressions communicated in letters of recommendation, or on personal interviews with candidates for admission to selective programs? These will simply reinforce existing biases. How about eliminating grades and basing admission to a selective institution, say MIT, on the basis of a lottery? Wouldn't educating such an enormous range of students, from illiterate students to prodigies, make the instructional task virtually impossible to accomplish? Wouldn't public perception that a student's admission depended not only partially, but *entirely* on luck undermine the credibility of the institution itself? Although I do not favour efforts to

eliminate assessments of student achievement, that does not mean that the assessment process cannot be improved, and, indeed, in the United States most selective colleges and universities rely on a portfolio of evidence before deciding what strengths an applicant has to offer.

I think the hope lies not in abandoning the attempt to measure student academic achievement, but in creating a society in which academic achievement is less fateful. Unfortunately, this is a project that, itself, depends on an educated citizenry, for those who believe – often correctly – that the structures of society are rigged against them must find means to press their claims in credible ways. That is to say, they must persuade rather than simply overpower the more advantaged among us.

I daresay that no one who read my first paragraph, noting both psychological and academic costs of the pandemic, was surprised at my mentioning both in the same paragraph. Why not? Because the pandemic has brought home the fact that a single cause, like a highly contagious virus, can and does produce myriad effects, all of which we need to take account of. And yet, moving away from the assessment of individual students to the evaluation of programs and policies in education, evaluations performed by highly trained researchers, I am struck by their almost exclusive focus on academic achievement as indicated by scores on one or a few standardized tests (e.g., Hanushek et al., 2007; Hoxby, 2003; What Works Clearinghouse, 2007). These same researchers, claiming to emulate their colleagues in medical school, have designated the randomized clinical trial as the gold standard, yet it is clear to me that by ignoring side effects, their evaluations fall far short of those performed in the medical context.

Why Side Effects?

Why must evaluations track side effects? The answer for medicine as for education is the same. Jerry Avorn (2005) articulates it succinctly in his book on drug evaluation:

Any molecule clever enough to influence one chemical pathway for the better is also likely to have other effects in the complex biological soup of the cytoplasm (p. 71).

Most medical interventions produce myriad side effects, but some common ones (e.g., fatigue) are deemed relatively insignificant when measured against therapeutic or preventive effectiveness. On the other hand, side effects such as heightened risk of stroke or heart attack become significant because they put the patient's life at risk and therefore challenge a positive evaluation of a drug based on efficacy alone. Just recently the US Centers for Disease Control (CDC) and the Federal Drug Administration (FDA) recommended a pause in administration of Johnson and Johnson's COVID-19 vaccine due to six cases of unusual blood clots in over six million people who had been vaccinated. There is also the well-documented excessive use of opioids, which relieve pain but foster addiction. So drug efficacy alone is not sufficient; it must be balanced against the risk of deleterious side effects – “the efficacy/safety ratio” – and perhaps also against cost – “cost-effectiveness.” Let me note two further points: 1) Whether an outcome is labelled as the intended effect or a “side” effect depends *not* on the drug but on the intention of the prescriber. Aspirin may be taken for a headache or to assist circulation. Which one is considered the side effect depends on the reason for which one is taking it. 2) Determining the ratio of benefits to risks is far from straightforward; it is often contested, even among experts, as the current debate over booster shots to protect against COVID-19 or the continuing debate over breast cancer therapies attests.

In the educational context, a claim corresponding to Avorn's was succinctly formulated over 80 years ago by John Dewey:

Perhaps the greatest of all pedagogical fallacies is the notion that a person learns only the particular thing he is studying at the time. Collateral learning in the way of formation of enduring attitudes, of likes and dislikes, may be and often is much more important than the spelling lesson or lesson in geography or history that is learned. For these attitudes are fundamentally what count in the future. The most important attitude that can be formed is that of desire to go on learning. (1938/1973, p. 48)

Collateral learning is produced by every educational intervention. Yet Dewey's analysis points to an important *disanalogy* between the medical and educational settings that must be noted before continuing: it is rare in medicine, but not in education, for collateral effects to be *beneficial*. Health professionals are, therefore, focused primarily on avoiding adverse effects. Educators, by contrast, seek *both* to avoid adverse effects and to foster beneficial ones.

Which Side Effects?

The problem in education is not the absence of “collateral learning” but its abundance. Among the dispositions likely to be affected (for good or ill) are self-awareness, a readiness to help others, tolerance, assiduousness, and conformity. Educational experiences are hardly ever either life threatening or life saving, so how do we identify the collateral learning that merits tracking? I claim that evaluators are obliged to assess that collateral learning which suffices to call into question a verdict based on academic achievement alone. I believe that one kind of collateral learning unequivocally meets this requirement – precisely that identified by Dewey, namely motivation to continue learning.

Dewey is not the only one to highlight the importance of such a motivation. In a 1976 article in *Review of Educational Research* devoted to the topic (which, surprisingly, fails to reference Dewey), psychologist Martin Maehr, noting its neglect, tried to press its importance upon the education research community. Apparently his message went unheeded. Following Maehr, I will hereafter refer to this “continuing motivation” to learn as CM (1976, p. 445.)¹

Let me be clear here: I am not (nor is Dewey or Maehr) claiming that a reduction in CM must always outweigh a gain in academic achievement. What I am committed to is the claim that *evidence concerning the strengthening or weakening of CM is always relevant to, and may either overturn or confirm a verdict based solely on, achievement measures*. Note that taking CM into account can challenge an evaluation based solely on efficacy from two directions. A program that generated lower achievement might be deemed worthier through producing gains in CM; a program that generated higher achievement might be deemed less worthy based on a demonstrated loss in CM.

There are two additional reasons to focus on CM. First, of all the side effects produced by an educational program or policy, CM is the one that is undeniably at the core of the academic mission of the school, a mission that almost all consider the *raison d'être* of the school in the first place. Second, one reason someone might deny the need to consider CM is a belief that achievement gains will, as a matter of psychological fact, be accompanied by gains in CM. The reasoning here is plausible, but one of the aims of evidence-based education is precisely the desire to seek evidence rather than depend on plausible reasoning alone. Despite its plausibility, this reasoning has been challenged by a substantial body of psychological research. Ever since Edward Deci's classic 1971 article, which showed that a focus on external rewards can weaken intrinsic motivation, the role of different reward structures on various kinds and levels of motivation has been hotly debated (Eisenberger & Cameron, 1996, Ryan & Deci, 2000). In the article cited above, Maehr (1976) notes that

authoritarian control, the use of bribes, rewards, and threats may have some immediate, desirable effects on performance, but result in unfortunate consequences as far as the development of CM is concerned (p. 446).

For example, many so-called “high-performing” charters have adopted a harsh disciplinary stance borrowed from “zero tolerance” policing, which has led to high suspension and attrition rates, especially among African American students, students learning English, and students with disabilities (Rizga, 2016.)

¹ Some might legitimately consider CM to be the intended effect and academic achievement a beneficial side effect. Either stance is defensible.

What is the impact of this kind of regime on the CM of students in these schools, both those who drop out and those who stay? We don't know, but should we not try to find out?

Assessing Continuing Motivation to Learn

How might the contributions of an educational program or structure to the development of CM be determined? The kind of study we need would track student *choices* subsequent to particular experiences. Consider the question of whether to issue grades of “incomplete” (rather than record failing grades) on students' report cards, as was done by about 20% of schools during the pandemic, according to the Rand report (Kaufman et al., 2021, p. 35). Follow-up studies could compare students graded in these two different ways, focusing not simply on their subsequent academic performance, but also on reasonable and available proxies for CM: attendance, homework completion, participation in class, etc. It is possible that students who received Fs lost CM and gave up. It is also possible that they encountered the failing grade as a wake-up call to take schoolwork more seriously. A carefully designed study can help us find out which is the case.

Can we devise experiments to elicit changes in CM before students react, through choices and actions, according to their natural dispositions? Imagine a randomized trial comparing an experimental literature program with a more conventional one for elementary school students. At the end of the school year, children are offered at no charge a set of six Caldecott Medal-winning books,² but in order to receive them they must first promise to write a book report, following a prescribed format, over the summer for two of the books. They are given the summer to perform the task. Their willingness to sign the contract will tell us something about whether the program has stirred their interest in reading, while the proportion of students who complete the book report will tell us whether the motivation was strong enough for them to complete the task.

Meeting Objections to the Argument

In concluding, let me focus on two objections that might be raised to the foregoing argument:

1. *While it is true that CM is important, the school is by no means the only contributor, perhaps not even the main contributor, to it. Therefore evaluations may safely ignore it, focusing on academic achievement alone.*

It is true that many dispositions, including CM, are affected by experiences both in and out of school, and it is also true that the impact of schooling on those dispositions may be limited, but the conclusion does not follow for two reasons. First, in the absence of empirical study, the magnitude of the school's contribution is simply not known. Secondly, although the family may enhance or undermine some general motivation to learn about topics that are dear to that particular family, be it politics or baseball, it is doubtful that most families in normal times would influence a child's CM to pursue specific academic subjects. But this might be different during school closures.

2. *Granted it would be desirable to assess CM, but assessing achievement is difficult enough, as the debate on the impact of voucher and charter schools illustrates. Adding a requirement to consider CM as collateral learning will only compound the problem of reaching a clear consensus on any program or policy.*

Just *because* the advantages of most educational programs and policies, measured by test scores alone, are likely to be rather modest, the gain or loss in CM may well tip the balance in favour of or against the

² The Caldecott Medal is awarded to the “most distinguished American picture book for children.”

program or policy under review. Remember that drug evaluation would also be much simpler if efficacy alone were considered.

We are now learning, sometimes to our sorrow, that not only must medical drugs and procedures be efficacious for the conditions they address, but that such efficacy must outweigh any adverse effects they cause; so, likewise, evaluations of programs and policies designed to boost student achievement would be irresponsible if they failed to take losses as well as gains in CM into account. Let us not forget that there is an adverse side effect of the consistent focus on test scores to the exclusion of other outcomes: the “collateral learning” fades from our zone of awareness and we become guilty of the “greatest of all pedagogical fallacies.” Not only do we not bother to find out whether we are producing students with raised or lowered motivation to continue their learning, we no longer think it matters. The pandemic will have taught many of us that society cannot afford to ignore either academic achievement or CM. Let us hope the educational research community will have learned the same lesson.

References

- Avorn, J. (2005). *Powerful medicines: The benefits, risks, and costs of prescription drugs*. Rev. ed. New York: Vintage Books.
- Deci, E. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105–115.
- Dewey, J. (1973). *Experience and education*. New York: Collier Books. (Original work published in 1938).
- Eisenberger, R. & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, 51, 1,153–1,166
- Hanushek, E. A., Kain, J. F., Rivkin, S., & Branch, G. F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics*, 91, 823–848.
- Hoxby, C. M. (2003). School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review*, 10, 11–67.
- Kaufman, J. H., Diliberti, M. K., Hunter, G. P., Snoke, J., Grant, D., Setodji, C. M., & Young, C. J. (2021). COVID-19 and the state of k–12 schools: Results and technical documentation from the spring 2021 American educator panels COVID-19 surveys. Santa Monica, CA: Rand Corporation. https://www.rand.org/pubs/research_reports/RRA168-7.html
- Maehr, M. (1976). Continuing motivation: An analysis of a seldom considered educational outcome. *Review of Educational Research*, 46, 443–462.
- Panchal, N., Kamal, R., Cox, C., Garfield, R., & Chidambaram, P. (2021, May 26). Mental health and substance use considerations among children during the COVID-19 pandemic. KFF issue brief. <https://www.kff.org/coronavirus-covid-19/issue-brief/mental-health-and-substance-use-considerations-among-children-during-the-covid-19-pandemic>
- Rizga, K. (2016, March 18). The disturbing reason why some charter schools may have higher test scores. *Mother Jones*. <https://www.motherjones.com/politics/2016/03/charter-schools-suspend-more-black-students-disabilities-test-scores>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- What Works Clearinghouse. (2007). WWC intervention report: Knowledge is power program (KIPP). Washington: Institute of Education Sciences.

About the Author

Francis Schrag retired after a long career in the Departments of Educational Policy Studies and Philosophy at the University of Wisconsin, Madison. He published two books and numerous articles in both education and philosophy journals.