Like other people, educators often hold theories about how the world works, or how one ought to act, that are never named, never checked for accuracy, never even consciously recognized. One of the most popular of these theories is a very appealing blend of pragmatism and relativism that might be called “the more, the merrier.” People subscribing to this view tend to dismiss arguments that a given educational practice is bad news and ought to be replaced by another. “Why not do both?” they ask. “No reason to throw anything out of your toolbox. Use everything that works.”

But what if something that works to accomplish one goal ends up impeding another? And what if two very different strategies are inversely related, such that they work at cross purposes? As it happens, converging evidence from different educational arenas tends to support exactly these concerns. Particularly when practices that might be called, for lack of better labels, progressive and traditional are used at the same time, the latter often has the effect of undermining the former.

Example 1 comes from the world of math instruction. A few years back, a researcher named Michelle Perry published a
study in the journal *Cognitive Development* that looked at different ways of teaching children the concept of equivalence, as expressed in problems such as “4 + 6 + 9 = ___ + 9.” Fourth and fifth graders, none of whom knew how to solve such problems, were divided into two groups. Some were taught the underlying principle (“The goal of a problem like this is to find …”), while others were given step-by-step instructions (“Add up all the numbers on the left side, and then subtract the number on the right side”).

Both approaches were effective at helping students solve problems just like the initial one. Consistent with other research, however, the principle-based approach was much better at helping them transfer their knowledge to a slightly different kind of problem—for example, multiplying and dividing numbers to reach equivalence. Direct instruction of a technique for getting the right answer produced shallow learning.

But why not do both? What if students were taught the procedure *and* the principle? Here’s where it gets interesting. Regardless of the order in which these two kinds of instruction were presented, students who were taught both ways didn’t do any better on the transfer problems than did those who were taught only the procedure—which means they did far worse than students who were taught only the principle. Teaching for understanding didn’t offset the destructive effects of telling them how to get the answer. *Any* step-by-step instruction in how to solve such problems put learners at a disadvantage; the absence of such instruction was required for them to understand.

Example 2 has to do with how learning is evaluated. In a study that appeared in the *British Journal of Educational Psychology*, Ruth Butler took 5th and 6th graders, including both high- and low-achieving students, and asked them to work on some word-construction and creative-thinking tasks. One-third of them then received feedback in narrative form, one-
third received grades for their performance, and one-third received both comments and grades.

The first finding: Irrespective of how well they had been doing in school, students were subsequently less successful at the tasks, and also reported less interest in those tasks, if they received a grade rather than narrative feedback. Other research has produced the same result: Grades almost always have a detrimental effect on how well students learn and how interested they are in the topic they’re learning.

But because Ms. Butler had thought to include a third experimental condition—grades plus comments—she was able to document that the negative effects of grading, on both performance and interest, were not mitigated by the addition of a comment. In fact, with the task that required more original thinking, the students’ performance was highest with comments, lower with grades, and lowest of all with both. These differences were all statistically significant, and they applied to high- and low-achieving students alike. As in Michelle Perry’s math study, the more traditional practice not only didn’t help, but actually wiped out the positive effects of the alternative strategy.

One recalls the bit of folk wisdom—confirmed by generations of farmers and grocers—warning that a rotten apple can spoil a barrel full of good apples. It would be pushing things to postulate a kind of educational ethylene released by traditional classroom practices, analogous to the gas given off by bad fruit. But it does seem that the quest for optimal results may sometimes require us to abandon certain practices rather than simply piling other, better practices on top of them.

* *

In other instances, too, the rotten-apple theory offers a better fit with educational reality than does “the more, the
merrier.” Consider schools that try to have it both ways: They work with students who act inappropriately, perhaps even spending time to promote conflict-resolution strategies—but they still haven’t let go of heavy-handed policies that amount to doing things to students to get compliance. On the one hand: “We’re a caring community, committed to solving problems together.” On the other hand: “If you do something that displeases us (the people with the power), we’ll make you suffer to teach you a lesson.”

What might explain these mixed messages? Sometimes a school is in transition, grasping for something better but still holding on to old-fashioned control until everyone becomes sufficiently confident about the new approach to let go of the old. Sometimes a theory even more optimistic than “the more, the merrier” is at work: an “antidote” model that assumes the bad will be detoxified by the good. I haven’t seen any hard data one way or the other on this question, but plenty of anecdotal evidence suggests that some schools wind up taking away with one hand what they’ve given with the other. A peer-mediation program is nice, but its potential to do good is limited if kids are still subject to detentions, suspensions, rewards for obedience, and so on. As a principal in Connecticut observed, after describing her school’s struggle to create a more positive climate, “Our original goals were to control student behavior and build community, but along the way we learned that these are conflicting goals.” Only when the “doing to” is gone can the “working with” really begin to make some headway.

That smell of good apples going bad also issues from classrooms that try to combine collaboration and competition—for example, by putting students into groups but then setting the groups against one another. The reason for cooperative learning, students infer, is to defeat another bunch of students learning together. Cooperation becomes merely instrumental, the goal being to triumph over others.
Or consider a teacher who does all the right things to help kids love reading: surrounds them with good books and offers plenty of time to read them; gives kids choices about what to read and how to respond to what they’ve read; teaches them to read from the beginning through rich stories and other authentic material, with a focus on meaning rather than just on decoding skills. Sometimes, however, those ingredients of literacy are soured by the simultaneous use of reading incentives—either home-grown schemes or slick prefabricated programs (bought with precious book-acquisition funds)—that lead children to regard reading as a tedious prerequisite to receiving points and prizes. It’s hard to treat kids like budding bibliophiles when they’re also being treated like pets.

Underlying this last example, as well as Ruth Butler’s grading study and perhaps even the tension between problem-solving and discipline, is the deeper issue of motivation to learn. Or maybe we should say motivations to learn, because the point is that there are qualitatively different kinds. One of psychology’s most robust findings is that extrinsic motivation (doing something in order to receive a reward or avoid a punishment) is completely different from—and often inversely related to—intrinsic motivation (doing something for its own sake). The more we offer rewards to “motivate” people, the more they tend to lose interest in whatever they had to do to get the reward.

Some behaviorists have tried to challenge the growing evidence supporting that contention, but the latest major research review—see Psychological Bulletin, vol. 125 (1999): 627-68—dispels any lingering doubt about a finding that has by now held up across genders, ages, cultures, settings, and tasks: Two kinds of motivation simply are not better than one. Rather, one (extrinsic) is corrosive of the other (intrinsic)—and intrinsic is the one that counts. To make a difference, therefore, we have to subtract grades, not just
add a narrative report. We have to eliminate incentives, not just promote literacy. We have to remove coercive discipline policies, not just build a caring community.

* These days, with our attention riveted on the Tougher Standards version of school reform as on a slow-motion train wreck, we may, if we look very carefully, notice another illustration of the rotten-apple phenomenon playing out before our eyes. Top-down demands to raise scores on bad tests are terrible and ought to be vigorously opposed. But what about top-down demands to raise scores on reasonably good tests? What happens when states offer performance-based assessments, but in the context of “accountability” systems—basically, extrinsic pressure—to improve the results?

In a word, the former are destroyed by the latter. Exhibit A is the Kentucky Education Reform Act, rolled out in the early 1990s, which proposed to let students show what they understood rather than just memorizing facts and bubbling in ovals. Unfortunately, their performance triggered a series of rewards and penalties for educators, and schools quickly became pressure cookers. With so much riding on the outcome, technical concerns about reliability came to overshadow pedagogical concerns about improving learning.

Before the decade was out, the best features of the experiment had been dismantled, with conventional tests replacing richer measures. “High-stakes accountability and performance assessment are based on conflicting principles,” as Ken Jones and Betty Lou Whitford observed in their summary of the state’s reform. “One encourages conformity to externally imposed standards, while the other grows out of emergent interaction between teachers and students.”

Exhibit B is the Maryland State Performance Assessment Program, or MSPAP, a system begun around the same time as
Kentucky’s that has more recently met the same sad fate. It featured open-ended questions and authentic tasks to measure critical thinking, but it, too, was married to high stakes: Schools were publicly ranked, with bonuses for the high scorers and humiliation and threats for the low. Again, the quality of the assessment couldn’t protect students and teachers from the toxic effects of what now passes for “accountability”: The curriculum was narrowed to focus on MSPAP questions (for example, more structured writing, less creative writing), students had to memorize catchy formulas for producing high-scoring essays, and schools were set against each other in a mutually destructive competition. High-stakes meant high stress for high- and low-performing schools alike.

The death of the MSPAP had other causes, too: relentless opposition from conservatives (whose counterparts in California and Arizona had also succeeded in halting short-lived experiments with authenticity); pressure to chart the results of individual students, rather than sample their performance so as to monitor schools; and concerns about reliability and errors in scoring prompted by lower scores than expected in affluent areas this past spring.

These factors aside, though, there are two central lessons to be drawn from Maryland and Kentucky:

1. Even when the assessment is performance-based, teaching to the test is (a) possible, (b) undesirable, and (c) done pervasively (indeed, frantically).
2. Analogous to the economic principle known as Gresham’s Law, bad tests will drive out good tests in a high-stakes environment. The current accountability fad—which was launched for political, not educational, reasons—inexorably dumbs down assessment. It leaves us with the sort of conventional standardized tests that are more consistent with the purposes of rating and ranking, bribing and threatening.
Then again, we may be witnessing something that transcends the challenges of assessment, a macro echo of a phenomenon confirmed at the micro level: The bad stuff has to be eliminated for the good stuff to work.

Addendum: The basic idea described here was replicated in a 2004 Dutch study in which college students were given an article to read about recycling. Some were told that it would help them contribute to a healthy environment (an intrinsic goal); some were told it would help them save money (an extrinsic goal); some were told both. Those students led to focus on the intrinsic goal ended up more interested in what they were reading and also understood the concepts better than those led to focus on the extrinsic goal — and also better than those given both goals. “It appears that adding an extrinsic goal to an already present intrinsic goal directs people’s attention away from the learning task to external indicators of worth…which in turn forestalls their performance…and undermines their persistence,” the authors concluded.

To be notified whenever a new article or blog is posted on this site, please enter your e-mail address at www.alfiekoohn.org/sign-up.