American Enterprise Institute for Public Policy Research

AEI

No. 1 • January 2009

Education Outlook

The New Stupid: Limitations of Data-Driven Education Reform

By Frederick M. Hess

A decade ago, it was disconcertingly easy to find education leaders who dismissed student achievement data and systematic research as having only limited utility when it came to improving schools or school systems. Today, we have come full circle. Educators have made great strides in using data, and it is hard to attend an education conference or read an education magazine without encountering broad claims for data-driven education reform. But danger lies ahead for those who misunderstand what data can and cannot do.

Phrases such as "data-driven decision-making" and "research-based practice" can readily morph into convenient buzzwords that stand in for careful thought, obscure rather than clarify, serve as dressed-up rationales for the same old fads, or justify incoherent proposals. Because few educators today are inclined to denounce data, there has been an unfortunate tendency to embrace glib new solutions rather than ask the simple question: what exactly does it mean to use data or research to inform decisions?[1]

Three Elements of the New Stupid

Today's enthusiastic embrace of data has waltzed us directly from a petulant resistance to performance measures to a reflexive and unsophisticated reliance on a few simple metrics—namely, graduation rates, expenditures, and the reading and math test scores of students in grades three through eight. The result has been a pirouette from one troubling mind-set to another; we have quickly pivoted from the "old stupid" to the "new stupid." The new stupid has three key elements.

Frederick M. Hess (rhess@aei.org) is a resident scholar and the director of education policy studies at AEI. A version of this article appeared in *Educational Leadership* in December 2008.

Using Data in Half-Baked Ways. I first encountered the inclination to energetically misuse data a few years ago while giving a presentation to a group of aspiring superintendents. They were passionate, eager to employ research and make data-driven decisions, and committed to leaving no child behind. We had clearly left the old stupid in the rearview mirror. New grounds for concern emerged, however, as we discussed teacher assignments and value-added assessment—the measure of academic gains that can be attributed to an individual teacher.

The group had recently read a research brief highlighting the effect of teachers on student achievement as well as the inequitable distribution of teachers within districts, with higher-income, higher-performing schools getting the first picks. The aspirants were fired up and ready to put this knowledge to use. To a roomful of nods, one declared, "Day one, we're going to start identifying those high-value-added teachers and moving them to the schools that aren't making [adequate yearly progress]."

Now, although I was generally sympathetic to the premise, the certainty of the stance provoked me to ask a series of questions: Can we be confident that teachers who are effective in their current classrooms would be equally effective

elsewhere? What effect would shifting teachers to different schools have on the likelihood that teachers would remain in the district? Are the measures in question good proxies for teacher quality? What steps might either encourage teachers to accept reassignment or improve recruiting for underserved schools?

My concern was not that the would-be superintendents lacked firm answers to these questions—that is natural even for veteran big-district superintendents who are able to lean on research and assessment departments. It was that they seemingly regarded such questions as distractions. One aspirant captured the mind-set perfectly when she said, "We need to act. We've got children who need help, and we know which teachers can help them."

At that moment, I glumly envisioned a new generation of superintendents shuffling teachers among schools—perhaps paying bonuses to do so—becoming frustrated at the disappointing results, puzzling over the departure of highly rated teachers, and wondering what had gone wrong. This is what it must have been like to listen to eager stock analysts in 1998 explain why some hot new Internet start-up was a sure thing while dismissing questions about strategy and execution as evidence that the stodgy questioners "just didn't get it."

Then as now, the key is not to retreat from data but to embrace it by asking hard questions, considering organizational realities, and contemplating unintended consequences. Absent sensible restraint, it is not difficult to envision a raft of poor judgments governing staffing, operations, and instruction—all in the name of data-driven decision-making.

Translating Research Simplistically. For two decades, advocates of class-size reduction have referenced the findings from the Student Teacher Achievement Ratio (STAR) project, a class-size experiment conducted in Tennessee in the late 1980s. Researchers found significant achievement gains for students in small kindergarten classes and additional gains in first grade, especially for black students. The results seemed to validate a crowd-pleasing reform and were famously embraced in California, where in 1996 legislators adopted a program to reduce class sizes that cost nearly $800 million in its first year and billions in its first decade. The dollars ultimately yielded disappointing results, however, with the only major evaluation (a joint American Institutes for Research and RAND study[2]) finding no effect on student achievement.

What happened? Policymakers ignored nuance and context. California encouraged districts to place students in classes of no more than twenty—but that class size was substantially larger than those for which STAR found benefits. Moreover, STAR was a pilot program serving a limited population, which minimized the need for new teachers. California's statewide effort created a voracious appetite for new educators, diluting teacher quality and encouraging well-off districts to strip-mine teachers from less affluent communities. The moral is that even policies or practices informed by rigorous research can prove ineffective if the translation is clumsy or ill-considered.

Even policies or practices informed by rigorous research can prove ineffective if the translation is clumsy or ill-considered.

When it comes to "research-based practice," the most vexing problem may be the failure to recognize the limits of what even rigorous scientific research can tell us. For instance, when testing new medical treatments, randomized field trials are the research design of choice because they can help establish cause and effect. Efforts to adopt this model in schooling, however, have been plagued by a flawed understanding of just how the model works in medicine and how, or if, it can be translated to education. The randomized field trial model, in which drugs or therapies are administered to individual patients under explicit protocols, is enormously helpful when recommending interventions for particular medical conditions. But it is far less useful when determining how much to pay nurses or how to hold hospitals accountable.

In education, curricular and pedagogical interventions can indeed be investigated through randomized field trials, with results that can serve as the basis for prescriptive practice. Even in these cases, however, there is a tendency for educators to be cavalier about the elements and execution of research-based practice. When medical research finds a certain drug regimen to be effective, doctors do not casually tinker with the formula. Yet, in areas like reading instruction, districts and schools routinely alter the sequencing and elements of a curriculum, while still touting their practices as research-based.

Meanwhile, when it comes to policy, officials must make tough decisions about governance, management, and compensation that cannot be examined under

controlled conditions and for which it is difficult to glean conclusive evidence. Although research can shed light on how policies play out and how context matters, studies of particular merit-pay or school-choice plans are unlikely to answer whether such policies "work"—largely because the particulars of each plan will prove crucial.

Giving Short Shrift to Management Data. School and district leaders have embraced student achievement data but have paid scant attention to collecting or using data that are more relevant to improving the performance of schools and school systems. The result is "data-driven" systems in which leaders give short shrift to the operations, hiring, and financial practices that are the backbone of any well-run organization and that are crucial to supporting educators.

Existing achievement data are of limited utility for management purposes. State tests tend to provide results that are too coarse to offer more than a snapshot of student and school performance, and few district data systems link student achievement metrics to teachers, practices, or programs in a way that can help determine what is working. More significant, successful public and private organizations monitor their operations extensively and intensively. FedEx and UPS know at any given time where millions of packages are across the United States and around the globe. Yet few districts know how long it takes to respond to an applicant for a teaching job, how frequently teachers use formative assessments, or how rapidly school requests for supplies are processed and fulfilled.

For all of our attention to testing and assessment, student achievement measures are largely irrelevant to judging the performance of many school district employees. It simply does not make sense to evaluate the performance of a payroll processor or human resources recruiter—or even a foreign language instructor—primarily on the basis of reading and math test scores for grades three through eight.

Just as hospitals employ large numbers of administrative and clinical personnel to support doctors and the military employs accountants, cooks, and lawyers to support its combat personnel, so schools have a "long tail" of support staff charged with ensuring that educators have the tools they need to be effective. Just as it makes more sense to judge the quality of army chefs on the quality of their kitchens and cuisines than on the outcome of combat operations, so it is more sensible to focus on how well district employees perform their prescribed tasks than on less direct measures of job performance. The tendency to focus casually on student achievement, especially given the testing system's heavy emphasis on reading and math, allows a large number of employees either to be excused from results-driven accountability or to be held accountable for activities over which they have no control. This undermines a performance mind-set and promises to erode confidence in management.

Ultimately, student achievement data alone yield only a "black box." They illustrate how students are faring but do not enable an organization to diagnose problems or manage improvement. It is as if a CEO's management dashboard consisted of only one item—the company stock's price.

Data-driven management should not simply identify effective teachers or struggling students but should also help render schools and school systems more supportive of effective teaching and learning. Doing so requires tracking an array of indicators, such as how long it takes books and materials to be shipped to classrooms, whether schools provide students with accurate and appropriate schedules in a timely fashion, how quickly assessment data are returned to schools, and how often the data are used. A system in which leaders possess that kind of data is far better equipped to boost school performance than one in which leaders have a palette of achievement data and little else.

Steering Clear of the New Stupid

There are at least four keys to avoiding the new stupid. First, educators should be wary of allowing data or research to substitute for good judgment. When presented with persuasive findings or promising new programs, it is still vital to ask the simple questions: What are the presumed benefits of adopting this program or reform? What are the costs? How confident are we that the promised results are replicable? What contextual factors might complicate projections? Data-driven decision-making does not simply require good data; it also requires good decisions.

Second, schools must actively seek out the kind of data they need as well as the achievement data external stakeholders need. Despite quantum leaps in state assessment systems and continuing investment in longitudinal data systems, school and district leaders are a long way from having the data they require. Creating the conditions for high-performing schools and systems requires operational metrics beyond student achievement. In

practice, there is a rarely acknowledged tension between collecting data with an eye toward external accountability (measurement of performance) and doing so for internal management (measurement for performance).

The data most useful to parents and policymakers focus on how well students and schools are doing; this is the kind of data required by the No Child Left Behind Act and collected by state accountability systems. Although enormously useful, these assessments have also exacerbated a tendency of school and district leaders to focus on the data they have rather than on the data they need.

Education leaders should not expect research to resolve thorny policy disputes over school choice or teacher pay any more than medical research has ended contentious debates over health insurance or tort reform.

Current conditions call to mind the parable of the drunken man crawling under the streetlight while searching for his keys. A good Samaritan stops to help; after minutes of searching, she asks, "Are you sure you dropped your keys here?" The man gestures toward the other end of the street, saying, "No, I dropped them down there—but the light's better over here." We must take care that readily available data on reading and math scores for grades three through eight and high school graduation rates—all of which provide useful information—do not become streetlights that distract more than they illuminate.

Third, we must understand the limitations of research as well as its uses. Especially in the crafting of policy, research should not dictate outcomes but should instead ensure that decisions are informed by the facts and insights that science can provide. Researchers can upend conventional wisdom, examine design features, and help gauge the effect of proposed measures. But education leaders should not expect research to resolve thorny policy disputes over school choice or teacher pay any more than medical research has ended contentious debates over health insurance or tort reform.

Finally, school systems should reward education leaders and administrators for pursuing more efficient ways to deliver services. Indeed, superintendents who use data to eliminate personnel or programs—even if these superintendents are successful and vindicated by the results—are often more likely to ignite political conflict than to reap professional rewards. So long as leaders are revered only for their success at consensus building and gathering stakeholder input, moving from the rhetorical embrace of data to truly data-driven decision-making will remain an elusive goal in many communities. This is especially true given state and federal statutes, salary schedules, and established policies that restrict the ability to redeploy resources and that make aggressive efforts to act on data and research exhausting and contentious. The result is a chicken-and-egg conundrum, in which officials have limited incentive to track managerial data given their limited ability to use it, yet the resulting vacuum makes it more difficult to argue that flexibility will be used in informed and appropriate ways.

Research and data are powerful tools. Used thoughtfully, they are dynamic levers for improving schools and schooling. In this new era, educators stand to benefit enormously from advances in research and data systems. Let us take care that hubris, faddism, and untamed enthusiasm do not render these gifts more hindrance than help.

Notes

1. Frederick M. Hess, ed., *When Research Matters: How Scholarship Influences Education Policy* (Cambridge, MA: Harvard Education Press, 2008).

2. George W. Bohrnstedt and Brian M. Stecher, eds., *What We Have Learned about Class Size Reduction in California* (Sacramento: California Department of Education, September 2002), available at www.classize.org/techreport/CSRYear4_final.pdf (accessed January 16, 2009).