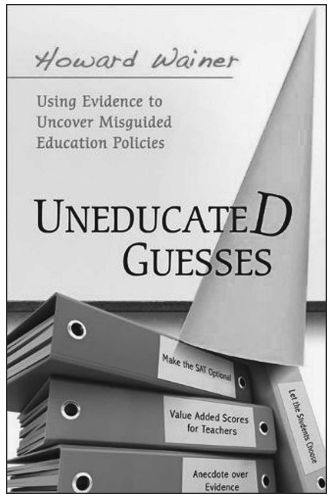
Uneducated Guesses: Using Evidence to Uncover Misguided Education Policies



Reviewed by John Ewing

Uneducated Guesses: Using Evidence to Uncover Misguided Education Policies
by Howard Wainer
Princeton University Press, 2011
US$24.95, 175 pages
ISBN-13: 978-0-691-14928-8

> *Uneducated Guesses* challenges everything our policymakers thought they knew about education and education reform.…In this explosive book, Howard Wainer uses statistical evidence to show why some of the most widely held beliefs in education today…are wrong.

So begins the description on the dust jacket, which promises that the author "exposes today's educational policies to the light of empirical evidence and offers solutions for fairer and more viable policies in the future." It claims this is an exposé that "no one who is concerned about seeing our children achieve their full potential can afford to ignore…" I wasn't looking forward to a book that claimed to solve all our educational problems.

It turned out, however, that the dust jacket was not merely hyperbole—it was just plain wrong. Rather than a book of policy, this was a collection of vignettes about the use of tests in education, mixed with occasional asides on elementary statistics, homely philosophy, and pleasant (if occasionally biting) wit. Not many books on statistics use references from the Bible in one chapter (Judges 12:4-6) and Indiana Jones in another, quote Hume and Kant alongside Feynman and Satchel Paige, or blend stories about Jaime Escalante and John Stuart Mill. This is a book with surprising charm, occasional insight, and many good stories, although not many "solutions for fairer and more viable policies." Ironically, the author *does* expose the way current educational policy is made, but not in the way the dust jacket suggests—more on this later.

Howard Wainer was for many years the principal research scientist at Educational Testing Service, and so it is not surprising that much of the material, especially in the first half of the book, concerns the use of standardized tests.

> *Question*: If you make submitting SAT scores optional for college admission, what happens?
>
> *Answer*: Those with high scores continue to submit SAT information; those with low scores do not. As a result, the value of SAT scores diminishes for screening applicants.
>
> *Question*: If instead of using a common entrance exam (an "aptitude test") we substitute scores on tests in a variety of subjects ("achievement tests"), what happens?
>
> *Answer*: It becomes more difficult to use tests to compare one applicant against another. One cannot easily judge whether being outstanding in history is as valuable as excelling in mathematics.
>
> *Question*: If we need to allocate a scarce resource (say, seats in an AP Calculus class) among a group of

John Ewing is president of Math for America. His email address is jewing@mathforamerica.org.

DOI: http://dx.doi.org/10.1090/noti874

> students, what is the most efficient way to do it?
>
> *Answer*: Use scores on a standardized aptitude test (PSAT), which turn out to be a good predictor of scores on standardized AP exams at the end of the course. (The surprise is that success in AP Psychology is better predicted than success in AP Calculus.)

None of these answers is a surprise, and the real surprise is that sensible people would have thought otherwise. In fact, this is the true import of the account. College admissions officers apparently thought that making SAT scores optional merely improved the admissions process without distorting it. A major reputable organization (the National Association for College Admission Counseling) advocated using sets of achievement tests in place of a standard aptitude test without realizing that this made it difficult to compare applicants. The public may be hoodwinked into believing a particular teacher has an exceptional gift when teaching advanced placement courses, when in fact the real gift turns out to be selecting the right students before the course begins. The author scoffs at such foolishness and counters with data, graphs, and elementary statistics.

Some of the vignettes are prosaic. Chapter 10 on ranking colleges says very little of substance about the rankings, commenting that the *US News and World Report* ranking "does seem to reflect generally the consensus of opinion about the quality of colleges and universities." The author includes a short discourse on the treatment of missing data in preference ranking, but there isn't much anyone could disagree with here. Some of the vignettes are slightly tedious. Chapter 8 tells the story of a third-grade teacher who had sixteen of her twenty-five students with a perfect score on a standardized mathematics test. She was accused of cheating, and indeed there was the possibility that a teacher's aide had inadvertently helped students. A young Ph.D. with only a course in "measurement" under his belt affirmed the accusation. Professional statisticians showed that the case was weak. A modestly interesting story, but too long in the telling.

All the vignettes are told with alacrity, wit, and fervor. The lessons on statistics are mostly presented deftly so that someone with little training in mathematics can easily comprehend them. The prose is attractive and clear.

The one chapter that touches on a current, sensitive, and hotly debated issue was disappointing. Chapter 9 deals with teacher evaluation using value-added modeling, and unlike the rest of the book, readers who are new to the ideas of statistics will find this hard to digest. It is one of the few places where equations appear. The author begins to explain the value-added model:

> The model itself begins simply by representing a student's test score in the first year, Y_1, as the sum of the district's average for that grade, subject and year, say μ_1, and the incremental contribution of the teacher, say θ_1, and systematic and unsystematic errors, say ϵ_1. When these pieces are put together we obtain a simple equation for the first year,
>
> $$y_1 = \mu_1 + \theta_1 + \epsilon_1,$$
>
> or
>
> $$\text{Student's score}(1) = \text{district average } (1) + \text{teacher effect } (1) + \text{ error } (1).$$

The expression for "value-added," he notes, is "statistically convenient" because there are "fewer parameters."

This is an explanation that is insufficiently precise for the mathematically sophisticated and overwhelming for everyone else. More troubling, however, is his analysis of the efficacy of the model. Again, the unsophisticated reader will be confused and likely will come away with the notion that with a bit more effort one could overcome these "technical" problems. Near the end of the chapter, the author confirms this when he writes: "Value-added assessment may yet help us in this task, but there are many challenges yet to overcome before these models are likely to help us …" We just need to work harder, he seems to say. The fact that the model often reflects noise (and little else) is never mentioned.

This brings me to the most troubling aspect of this otherwise charming book: The author is a true believer in "data-driven education". This belief, shared by policymakers and politicians alike, has become the driving force behind education reform, and it is deeply troubling.

The book begins and ends with a discourse on epistemology. How do we find out about the world? What is the basis of our beliefs? What information should guide our actions? How do we judge whether those actions are the right ones? These are fundamental questions in any area of human endeavor, but they are especially important in education, where understanding our goals, how we might achieve them, and whether we have succeeded are crucial to preparing future generations.

On the first page of the introduction, the author begins to lay out his answers to these questions

with a famous quote from Richard Feynman, made in the Messenger Lecturers at Cornell University in 1964, attempting to explain the scientific method:

> In general we look for a new law by the following process. First we guess it. Then we compute the consequences of the guess to see what would be implied if this law that we guessed is right. Then we compare the result of the computation to nature, with experiment or experience, compare it directly with observations, to see if it works. If it disagrees with experiment it is wrong. In that simple statement is the key to science. *It does not make any difference how beautiful your guess is. It does not make any difference how smart you are, who made the guess, or what his name is—if it disagrees with experiment it is wrong. That is all there is to it.* [Feynman, R. P., 1965. *The Character of Physical Law*, p. 156, Cambridge: MIT Press]

So far, so good (it's hard to disagree with Richard Feynman). But then the author engages in a rhetorical slight-of-hand when he writes:

> It was clear that Feynman placed evidence in an exalted position. It vetoed all else.

Notice how in Wainer's sentence the phrase "experiment or experience" is narrowed to the single word "evidence". Soon he replaces the word "evidence" with "data" and then restricts data even further:

> Evidence of success in contemporary education encompasses many things, but principal among them are test scores. When scores are high, we congratulate all involved. When they are low, we look to make changes. [p. 4]

As the book proceeds, this point is driven home in chapter after chapter. Evidence means test scores, success means higher scores, "misguided policies" means anything that is not supported by higher scores. If evidence truly "encompasses many things," those many things are missing from this account altogether.

The American philosopher Abraham Maslow is often quoted as saying: "If the only tool you have is a hammer, you tend to see every problem as a nail." Statisticians see evidence as data, because their discipline is the tool that is associated with data. For modern researchers in education, their hammer is statistics and their nails are test scores—nothing else matters, no other evidence is valued.

But education is much more than tests, and educational evidence is much more than scores. Education is about critical thinking, about developing taste, empathy, and values. Education is about learning to carry out complicated analyses of complex problems over extended periods of time. Education is about learning to learn, so that whatever limited accomplishments one has in school are amplified throughout one's life. Education is much, much more, and little of this is measured by test scores. Evidence requires looking at students over many years, not one, and it includes information about the ongoing accomplishments of students, teachers, and administrators. Evidence even includes anecdotes (which the author dismisses as unworthy), because examples and case studies are derived from experience too. Evidence is not "data related to a claim" [p. 148]; evidence is "experience related to a claim." This is much, much larger than a nail.

The epilogue contains a clear and unambiguous description of the author's vision of data-driven education reform, which begins with an attack on the "experts".

> Experience has taught us a great deal about what kinds of optimization methods work in complex systems and what kinds do not. An almost surefire path to failure is to convene a blue-ribbon committee with a title like "Education 2020" whose mandate is to ponder existing problems and come out with recommendations for the future system. It doesn't work because even all-stars aren't that smart. [p. 157]

While I share his skepticism of committees, this seems to dismiss education professionals altogether. To emphasize that we should avoid "expert opinion" in crafting education, the author cites another famous quote of Feynman from the same 1965 Messenger Lecture: "Science is the belief in the ignorance of the experts." But he has taken this out of context and consequently changed its meaning. The full quote has a quite different message:

> We have many studies in teaching, for example, in which people make observations, make lists, do statistics, and so on, but these do not thereby become established science, established knowledge. They are merely an imitative form of

> science…The result of this pseudoscientific imitation is to produce experts, which many of you are. [But] you teachers, who are really teaching children at the bottom of the heap, can maybe doubt the experts. As a matter of fact, I can also define science another way: Science is the belief in the ignorance of experts. [Feynman, 1965]

A master teacher with years of experience who is an acknowledged craftsman in the profession surely is more likely to solve problems and shape the future of education than an economist, statistician, or politician. Richard Feynman seems to agree.

The epilogue goes on to explain how education policy should be made.

> What does work [in dealing with complex systems] is the implementation of constant experimentation, in which small changes are made to the process. Then the effects of the changes are assessed. If the process improves, the size of the changes is increased and the outcome monitored. If the process gets worse, the changes are reversed and another variable is manipulated. Gradually, the entire complex process moves toward optimization. [p. 157]

"Constant experimentation"—with the goal of higher test scores! When they go up, do a little more experimenting on the same variable; when they go down, try manipulating another variable. The results are predictable: The education system becomes corrupted; teachers find ways to manipulate scores, either by focusing their students' experience only on the tests or by cheating; administrators and politicians fixate on scores as the only evidence of value—the ultimate goal of all education; and the true educational professionals become disgusted and find something more rewarding to do with their lives. Surely this is not optimal.

This otherwise charming book inadvertently tells us a great deal about the nature of data-driven education. The author's goal is to help people think clearly about applying statistical thinking to education reform, and to some extent he succeeds. But statistics is only one tool among many, expertise can be found in disciplines other than statistics, and evidence is not synonymous with data. Some things are not measured by data and are not well suited to statistical analysis. The beauty of a great painting, the magnificence of a great symphony, the power of great poetry cannot be captured by data. Similarly, many prosaic parts of life are not aptly judged by data and statistics alone. Great and inspired education is among them.