

Education Review

Reseñas Educativas



Resenhas Educativas

May 6, 2026

ISSN 1094-5296

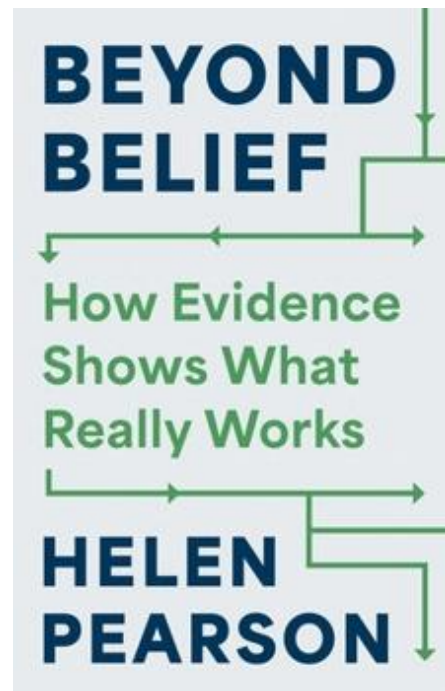
Pearson, H. (2026). *Beyond belief: How evidence shows what really works*. Princeton University Press.

368 pp.

ISBN: 978-0691207070

**Reviewed by Daniel H. Robinson
University of Texas at Arlington
United States**

As my frequent collaborator Howard Wainer has said, when faced with a difficult societal problem to solve, we must ask ourselves two questions. (1) Shall we use evidence? And if so, (2) what kinds of evidence are most relevant to the problem? Prior to the 1990s, the answer to the first question was mostly and sadly “No.” The term “evidence-based” prominently emerged in the early 1990s in medicine. This leaves one to wonder what was used before. Medical decisions relied heavily on the most senior doctor’s opinion – called “eminence-based medicine.” Only after the creation of the Cochrane Collaboration in 1993 did the medical community begin to rely regularly on syntheses of experimental evidence that came from published research studies.



Helen Pearson’s latest book is an important read for both those who embrace the latest anti-science movement and those who fear it. Very simply it is a page turner; it is hard to put down. A book calling for the use of evidence to evaluate what really works is beyond timely. It is indeed sad that such a book is needed. In the United States, we witness new examples of ridiculousness daily that make one wonder if what is being reported is part of a comedic farce rather than serious news. Sadly, this is not the case. The U.S. Secretary of Health and Human Services, Robert F. Kennedy Jr., spouts anti-vaccine nonsense. Due to people like RFK Jr., the US has witnessed a resurgence of measles cases that had long been eradicated.

Pearson tells a convincing story, using examples from a wide range of disciplines, of the importance of considering evidence in decision making. She smartly uses evidence to make the case for using evidence. No, she does not report

the results of experiments, but rather, she gives numerous examples where experiments revealed clear answers to questions that had long proved resistant to solution. It will be difficult for the reader to come away being anything other than what she calls a *randomista* – a fan of controlled experiments.

Pearson's telling of how medicine became evidence based is a fantastic story as are the stories from other fields that made similar monumental advances by synthesizing experimental studies. Accumulated evidence in the medical field allowed for delayed amputations, identified citrus fruits as a prevention for scurvy, streptomycin for tuberculosis, and led to rejection of Thalidomide for morning sickness. Experiments – specifically randomized controlled trials (RCTs) and meta-analyses – changed medical practice. Even non-RCT studies, when accumulated and studied carefully, can advance science. The Sudden Infant Death Syndrome (SIDS) epidemic of more than 60 years ago was at least partly due to parents relying on the most popular book at the time by Dr. Spock who encouraged them to lay babies on their stomachs to avoid asphyxiation due to vomiting. An RCT was not feasible nor ethical. Only after multiple case-control studies did the evidence emerge suggesting that such advice put babies at a higher risk for SIDS. It took until the 1990s for the medical community to finally urge parents to lay babies on their backs. Paying attention to evidence can save lives. As Pearson notes, the history of medicine is a story of practices that were once assumed to work but later found to be useless or harmful.

Pearson does a wonderful job of turning interviews into stories. Those stories begin with a problem that could not be solved without evidence. David Sackett used problem-based learning as a teaching strategy at McMaster University to teach medical students critical thinking. The students quickly learned that RCTs were better than observational studies and anecdotes. Sackett wanted to see if his new course taught doctors to be better. He designed an RCT with two hospitals that received the training with two control hospitals. It worked.

Sometimes RCTs aren't feasible or ethical. When that happens, it is best to do the next best thing. As with the case-control studies used to study SIDS causes, try to best simulate an RCT. Natural experiments can reveal causation. Evidence-based decisions are also possible for social policy. In 1986, Judith Gueron and the Manpower Demonstration Research Corporation (MDRC) ran RCTs to test Reagan's welfare-to-work policy. It had modest but positive effects. Similarly, the success of *Progresa*, a policy in Mexico designed to counter poverty, was confirmed with RCTs.

Pearson reveals an important limitation of policy research. Along with a need to know, sometimes there is a strong need to NOT know – especially if the people backing the policy fear that it will not show positive outcomes. RCTs serve as kryptonite to such ill-informed policies. Robert Boruch said that good evaluators should expect to get fired. Boruch helped form the Campbell Collaboration in 2000. Campbell is to social science research what Cochrane is to medical research. Unfortunately, synthesis research tends to receive much less funding than original research.

Long before the George Floyd murder in 2020, Minneapolis had challenges in policing. RCTs have contributed to evidence-based policing. Is it better to patrol all areas or just those with high crime? A 1980s RCT found that arrests lead to fewer repeated domestic violence incidents than mediation and counseling. But that is too general. A closer examination revealed that arrests work better when the perpetrator is employed or lives in richer areas.

In the field of conservation, most practices were based on anecdotes or myths. RCTs revealed that bat bridges, counting tigers, flooding reed beds, and stream engineering were all bad ideas. The Master of Business Administration (MBA) is one of the most popular degrees post-baccalaureate. Despite evidence-based management, some MBA courses teach anecdotal case studies rather than evidence-based strategies. In the absence of such evidence, companies hire consulting firms that may or may not use evidence. RCTs revealed that performance-related pay does not work as hoped for companies, and yet, many companies still use it. Pearson suggests that trying to effect change in business is futile. Instead, it is better to share learning modules on evidence-based management with universities who teach business students. Education would do well to follow this advice: Require evidence-based teaching courses as part of teacher preparation.

Speaking of education, Pearson finally gets around to discussing it in Chapter 7. Sadly, like medicine, the history of education is also a story of practices that were once assumed to work but later found to be useless or harmful. Shortly after George W. Bush was elected as president in 2000, education policy and practice seemed headed towards becoming more evidence based. The What Works Clearinghouse (WWC) was created in 2002 as a digital library of evidence-based educational research. Russ Whitehurst became the head of the newly formed Institute of Educational Science (IES). The IES provided grant funding for RCTs in education! The goal was to discover what works when, where, why, and for whom. And most importantly, one wants to know if the effect is big enough to make a meaningful difference, justify the expense, and so forth. Pearson is hopeful that education can witness similar advances as experienced in medicine, social policy, and conservation.

This hope is dampened by reality. In 2026, we see a different story. The IES has been gutted by the Trump Administration and perhaps will soon be dismantled as is the plan for the Department of Education. But did the IES do any good since its creation? What are the lasting effects? The WWC has been controversial because many have questioned its utility. Alan Schoenfeld and others have referred to the WWC as “What doesn’t work” due to the failures to provide clear consensus on many practices. Education simply does not have a plethora of high quality RCTs as one finds in medicine. And unlike other fields, the situation is getting worse rather than better. Yes, we may have 10 new meta-analyses in education being published every day. But the studies that are included in the meta-analyses are typically not RCTs. We are still falling short of Sir Iain Chalmers’s wish that we cumulate evidence scientifically.

Most treatments in medicine do not have dramatic effects. This is even more the case with education. The best interventions evaluated using the most rigorous RCTs typically reveal modest benefits. And they are not likely to be replicated for two reasons. First, there is the file drawer problem that continues to haunt education.

Most empirical journals are biased toward only publishing studies that find statistically significant effects. Those that find no effects are left in a file drawer. The researcher keeps trying and publishes one that finally hits. But the typical effect one can expect is minuscule to nothing. Hence, the file drawer problem.

I have a photo of my best bowling score from more than 10 years ago. I rolled a 209. Not surprisingly, I have no photos of any of my other scores. If I were to go out and bowl a game today, would I expect to roll over 200? Absolutely not. Like using a portfolio as an assessment where only one's best work is shown, journals present only a fraction of studies that likely do not represent reality. This is similar to using superscores in admission testing. Students are allowed multiple attempts and only their highest score counts. But the highest score is a biased estimate of their true score (Wainer & Robinson, 2025). A good meta-analysis accounts for effect size differences between published and unpublished studies. And there are several methods to adjust the estimated effects based on the file drawer problem. But these methods rely on heroic assumptions about the number of null or negative studies that are unknown and remain in the file drawer.

Second, the person who designed the intervention is usually the same one who is evaluating the intervention. This is a clear conflict of interest. Few want it known that they designed a weak intervention that failed to reveal a positive effect. Many researchers are funded by grants. Continued funding depends on success. Tenure and promotion also depend on publications and journals favor studies that find effects. Imagine if I was up for an award and I was also the person who decides who gets it. Such it is with having interventionists also serve as their own evaluator. As Pearson notes, the UK counterpart of the WWC is the Education Endowment Foundation (EEF). The people who created the EEF were smart enough to require independent outside evaluators for funded educational research. Someone proposes a new intervention and then a team with no connection to the person is responsible for conducting the RCT. We need to follow this requirement in the United States, especially given the recent replicability crisis in psychology. Rather than seeing one thousand new interventions each year, perhaps we would evaluate only a handful of the most promising to determine if they would scale up from laboratory to classrooms.

Can we easily turn things around in education? Unfortunately, all current signs point to pessimism. In my field of educational psychology, the percentage of empirical articles that employ RCTs has decreased from about 50% back in the early 1980s to 20% today. Instead, observational (mostly correlational) methods dominate the field. Most troublesome is the fact that researchers are providing recommendations for practice based on observational research about two-thirds of the time (Brady et al., 2023). In Chapter 9, I found an interesting quotation: "Most evidence comes from observational studies in which researchers look for correlations between measures of adolescent media use and mental health." Replace the last two terms with anything else and you get a sense of the typical research article appearing in educational psychology journals today. As other fields have become enlightened to the importance of RCTs in the last 30 years, education has turned its back on them. Abhijit Banerjee commented that it would be hard to find a

development economist under 50 who hasn't done an RCT. By contrast, it would be hard to find an educational researcher under 50 who has done one.

David Sackett developed a course at McMaster University called *Critical Appraisal of the Clinical Literature*. The purpose was to teach students how to separate the research wheat from the research chaff. I teach a similar course called *Evaluating and Debunking Educational Interventions*. My students are mostly teachers and report feeling betrayed when they learn that teaching to one's learning style and developing a growth mindset aren't quite what they were promised to be. A glaring omission in teacher education programs is teaching future teachers critical thinking and how to identify educational quackery. As Pearson notes, medical schools in the 1960s were great at teaching facts about biology but failed to teach doctors how to treat people. Schools of education are great at teaching theories but fail to teach teachers how well each approach works (and for whom). They don't do the latter because they simply don't know. Effects for educational interventions can be best estimated by randomly assigning students to either treatment or control groups.

Doctors prior to the 1990s and today's educators prefer to rely on their experience, discuss with colleagues, or defer to their intuitions rather than consult evidence to inform their practice (Lysenko et al., 2014). Gordon Guyatt, who coined the term "evidence-based medicine," realized that not everyone needs training in critical appraisal. Only a few need such training and then syntheses can be shared with everyone. So now we have the website UpToDate where doctors can get the latest on anything. Given the limitations of the WWC, we need an UpToDate for education.

The Obama Administration's Every Student Succeeds Act (ESSA) in 2015 introduced evidence standards for evaluating programs. Johns Hopkins University runs a free website called Evidence for ESSA (<https://www.evidenceforessa.org/>). Intrigued, I explored the site and quickly discovered that Reading Recovery was listed as having strong support. This is the same Reading Recovery that was mentioned in the podcast, "Sold a Story: How Teaching Kids to Read Went So Wrong." The intervention was under-evidenced and oversold. How trustworthy is a website that claims to synthesize evidence when such a glaring error occurs? The Education Endowment Foundation's Teaching and Learning Toolkit is a much better source for teachers who want to know quickly if something is evidence based (<https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit>).

I loved the "Ask for Evidence" campaign run by the group Sense about Science. They encourage people to challenge claims by asking for supporting evidence. They suggest the following message:

Dear Sir or Madam,

I was interested to [see/read in/hear] the claim [you/your organization/your company] made that [insert claim here]. Please can you send me the evidence you have to support this claim? I am asking as part of the Ask for Evidence campaign and will share the response I get publicly.

Yours sincerely . . .

My friend and collaborator Rob Bligh has sent dozens of similar emails to educational policymakers and teacher education faculty. The typical reply is no reply. Crickets. Most who claim to be education leaders have no evidence to support the nonsense they spew. They are what Pearson calls ZEBRAs (zero evidence but really arrogant). We must do better in using evidence to solve educational problems. Pearson's book is encouraging. Other fields are doing it. Why can't we?

References

- Brady, A. C., Griffin, M. M., Lewis, A. R., Fong, C. J., & Robinson, D. H. (2023). How scientific is educational psychology research? The increasing trend of squeezing causality and recommendations from non-intervention studies. *Educational Psychology Review*, 35, 36–37. <https://doi.org/10.1007/s10648-023-09759-9>
- Lysenko, L. V., Abrami, P. C., Bernard, R. M., Dagenais, C., & Janosz, M. (2014). Educational research in educational practice: Predictors of use. *Canadian Society for the Study of Education*, 37(2), 27. <https://www.jstor.org/stable/canajeducrevucan.37.2.06>
- Wainer, H., & Robinson, D. H. (2025). *Testing and the paradoxes of fairness*. Cambridge University Press. <https://doi.org/10.1017/9781009576857>

About the Reviewer

Daniel H. Robinson is professor and interim chair of the Department of Higher Education, Adult Learning, and Organizational Studies at the University of Texas at Arlington. He served as associate dean of research at UTA from 2020 to 2024. He received his PhD in educational psychology in 1993 from the University of Nebraska where he majored in both learning/cognition and statistics/research. He has taught at several universities including the University of Texas at Austin (1999-2012), and Colorado State University (2012-2015). He is the Specialty Chief Editor of *Frontiers in Psychology: Educational Psychology* and editor of *Monographs in the Psychology of Education: Child Behavior, Cognition, Development, and Learning*, Springer Publishing. He has previously served as editor of *Educational Psychology Review* (2006-2015), associate editor of the *Journal of Educational Psychology* (2014-2020), and as an editorial board member of nine journals. He is the co-author with Howard Wainer of *Testing and the Paradoxes of Fairness* (Cambridge, 2025) and co-editor (with Michael Mayrath and John Behrens) of *The Handbook of Generative Artificial Intelligence in Education: Integrating Research into Practice* (Springer, 2026). Dr. Robinson was a Fulbright Specialist Scholar at Victoria University, Wellington, New Zealand in 2011.



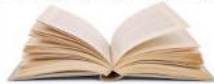
About the Author

Helen Pearson is an Honorary Professor of Practice at University College London, where she teaches science writing and journalism with the MSc in Science Communication. Most of her career has been built on a passion for science and writing. After a degree in natural sciences at the University of Cambridge, she earned a PhD in developmental genetics at the University of Edinburgh, after which she moved into science journalism. Dr. Pearson has been writing and editing for the journal *Nature* for over 20 years. For five years she was Chief Magazine Editor, leading the award-winning team of 60-70 journalists and editors that publish the magazine's journalism and opinion content. For this work, she won the Editor of the Year award from the Association of British Science Writers. She was named European Science Journalist of the Year in 2025. In 2026, she won a Knight Science Journalism Fellowship at MIT and will be in residence be in Boston from August 2026-May 2027.



Education Review

Reseñas Educativas



Resenhas Educativas



Education Review / *Reseñas Educativas* / *Resenhas Educativas* is supported by the Scholarly Communications Group at the Mary Lou Fulton College for Teaching and Learning Innovation, Arizona State University. Copyright is retained by the first or sole author, who grants right of first publication to the *Education Review*. Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and *Education Review*, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>. *Education Review* is free-to-read and free-to-publish.



Disclaimer: The views or opinions presented in book reviews are solely those of the author(s) and do not necessarily represent those of *Education Review*.