On education miracles in general (and those in Mississippi in particular)

Howard Wainer, **Irina Grabovsky** and **Daniel H. Robinson** explain why we should treat dramatic educational success stories in the USA with extreme caution

Matt Gush/shutterstock.com

The state moved from a 49th place ranking in 2013 to the top 20 in 2023. The latest 2024 scores revealed that Mississippi is now tied for 8th place among 53 US states and territories!

In 1748, famed Scot David Hume defined a miracle as a violation of a law of nature. He elaborated such a law as "a regularity of past experience projected by the mind to future cases". He argued that the evidence for a miracle is rarely sufficient to suspend rational belief because a closer look has always revealed that what was reported as a miracle was more likely false, resulting from misperception, mistransmission, or deception.[1] Hume's contemporary, Immanuel Kant, took a similar tone, but to avoid conflict with established religious beliefs he hedged his bets by explaining that miracles seemed to have been more frequent in the distant past but have become very rare recently. Our experience over the two centuries since their path-breaking work has only reinforced their caution in accepting the validity of any event whose very existence contradicts well-confirmed natural laws.

With Hume and Kant providing our epistemological background it was no surprise that we were sceptical when we read Noah Spencer's 2024 article[2] about "Mississippi's education miracle" which education economics expert Harry Anthony Patrinos[3] called a "model for global literacy reform". The results Spencer reported from his econometric model do seem to be miraculous – at least if you say them fast. Based on the National Assessment of Educational Progress (NAEP) fourth-grade literacy test scores, the state moved from a 49th-place ranking in 2013 to the top 20 in 2023. The latest 2024 scores revealed that Mississippi is now tied for 8th place among 53 US states and territories! Such a dramatic turnaround clearly marks a sharp deviation from what we expect given the laws of nature/education generated by a century of empirical experience. If the turnaround is indeed legitimate, then the "intervention" that is claimed to be the cause of the improvement, the Literacy-Based Promotion Act (LBPA), which started in 2013, should be seriously considered for implementation in other states.

A careful examination confirms that enthusiasm to emulate Mississippi should be tempered with scepticism. We add our humble voices to those of Hume and Kant that claims of miracles in the distant past are harder to dismiss because their evidence is blurred by the passage of time, but contemporary claims, when evidence is accessible, must be viewed with the scepticism justified by our knowledge of prior modern events whose very existence violates the claimed experience. Is our scepticism of Mississippi's recent improvement in NAEP performance justified?

We have seen several previous K–12 education "miracles" that turned out to be hoaxes. Five of them were in Houston, Atlanta, the District of Columbia, El Paso, and New Orleans.[4,5] In the first four, investigators found fraud. The people in charge (e.g., superintendents) cheated to give the impression of increased test scores. In Houston, the numbers of students who were categorised as "special education" were increased so their low test scores would not be included in the school's overall test scores.[6] In Atlanta, records were falsified. In the District of Columbia, high-school students graduated who should not have. And in El Paso, to inflate scores, Mexican transfer students, who typically scored lower, were prevented from taking the state-mandated tenth-grade achievement tests.

The New Orleans miracle was caused by a natural disaster. Hurricane Katrina tragically relocated about a third of the students who came from the poorest areas. Removing thousands of low scorers immediately raised the average test scores of the students who remained and it did so without increasing any student's individual score.

The Houston, El Paso, and New Orleans examples are of particular relevance to the latest Mississippi miracle. The improvement in the average performance of Mississippi's fourth-graders on NAEP was preceded by two key changes in their schooling in third grade. One was the *a priori* sensible idea of trying to improve classroom instruction by improved teacher training, instituting preschool, and a variety of other helpful actions. This was to be accomplished through the promise of an additional annual state expenditure of $15 million for the 134,376 students enrolled in kindergarten and up to third grade (in 2017). This provides a boost of about $111.63 of extra funding annually for each pupil. Comparing this amount to what are annual contemporary per pupil expenditures nationally, we have to agree that if such small expenditures can make a visible difference in student performance it truly is a miracle – a Mississippi version of St. John's loaves and fishes.

But it was the second component of the Mississippi Miracle, a new retention policy,

Howard Wainer is a statistician and author, and a former principal research scientist at the Educational Testing Service. He is the author of the books *Truth or Truthiness: Distinguishing Fact from Fiction by Learning to Think Like a Data Scientist* (Cambridge University Press, 2016), *A History of Data Visualization and Graphic Communication* (Harvard University Press, 2021) and, most recently, *Testing and the Paradoxes of Fairness* (Cambridge University Press, 2026).

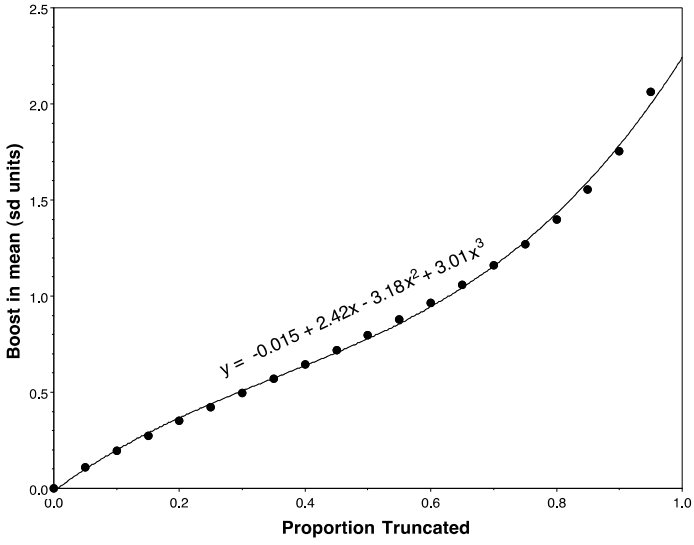

Figure 1: Curve showing how much the mean score will increase as a function of the proportion of the lowest scores truncated. Included is a third-degree polynomial that might be termed the "miracle worker". Insert (as x) the proportion to be truncated and it outputs the gain of the mean in the remaining data distribution. The details of how this is calculated are contained in the boxed technical appendix.

Table 1: Mississippi third-grade retention rates. Data is from *Mississippi Today* (tinyurl.com/mrxzpxwd)

	Retained	Total	Percent retained
2018–19	3,379	35,277	9.6%
2021–22	2,958	31,348	9.4%
2022–23	2,287	31,815	7.2%

Technical appendix

The curve in Figure 1 was calculated by assuming that reading scores in the population of 4th - graders are distributed as $N(\mu, \sigma^2)$ with the mean μ variance σ^2. Thus

$$\phi(t;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\}$$

is its probability density function.

Denote by $p(0 \le p < 1)$ the proportion of lowest-scoring students left behind to repeat the fourth grade. We want to compute the mean score of the remaining fourth-graders as a function of p. The mean score of this truncated distribution is equal to

$$\mu(p) = \frac{\sigma}{1-p}\phi(\Phi^{-1}(p)) + \mu,$$

where $\phi(t)$ is the probability density function of $N(0,1)$ and $\Phi(x)$ is its cumulative density function.

We define the relative boost function $rboost(p)$ of the mean scores after dropping the bottom $100p$% of the scores as the proportion of the standard deviation of the population scores:

$$rboost(p) = \frac{\mu(p)-\mu}{\sigma} = \frac{\phi(\Phi^{-1}(p)}{1-p}$$

perhaps inspired by New Orleans' Katrina disaster a decade earlier, that is likely to be the key to their success. Third-graders who fail to meet reading standards are forced to repeat the third grade. Prior to 2013, a higher percentage of third-graders moved on to the fourth grade and took the NAEP fourth-grade reading test. After 2013, only those students who did well enough in reading moved on to the fourth grade and took the test.

It is a fact of arithmetic that the mean score of any data set always increases if you delete some of the lowest scores (what is technically called "left truncation of the score distribution"). Those who choose to adopt such a policy need only consult the function shown in Figure 1 to learn what proportion of the data in the left tail needs to be truncated to obtain the amount of gain (in standard deviation units) that is desired.

As previously mentioned, the latest NAEP data for 2024 show even more impressive, "miraculous" results on the fourth-grade literacy test scores – a tie for 8th place. Strangely though, for the eighth-grade literacy test, the state's rank dropped to a tie for 42nd place! This should clear up any miracle illusions that may remain.

A sampling of typical third-grade retention rates is shown in Table 1. There are similar figures associated with earlier grades, but it is unknown (at least to us) how they combine to yield total retentions. These figures provide an aid for interpreting the function in Figure 1, but even rough estimates tell us that these retention rates portend an epiphenomenal increase in NAEP scores of at least 0.25 standard deviations.

We note that this result is unaffected by what constitutes the data distribution. Thus, you can just as easily increase the average student height or average family's wealth as you can NAEP score, so long as the distribution of data includes within it the value that is your goal. (Note that this works especially well for student height, for after retaining the shortest third-graders for an extra year they will likely be taller when they are measured again a year later. It would be nice if the same were true for students struggling in academic subjects.)

Figure 2 displays the mean fourth-grade NAEP literacy scores for the USA and Mississippi over the past 35 years. Scores gradually increased for both from 1990 to 2015. But then the scores began to decline nationally whereas they continued to increase for Mississippi. Why? It is hard to credit Mississippi's 2013 LBPA as the cause since there does not appear to be any change in Mississippi's continued improvement. Yet viewed in the context of the national decline, perhaps LBPA deserves some credit. But if we are to credit LBPA for the continued growth,

Irina Grabovsky, PhD, is a senior psychometrician at the National Board of Medical Examiners. Her research interests include test validity, measurement modelling, and methods for detecting irregular testing behaviour. Outside of work, she enjoys puzzles and reflecting on statistical paradoxes in education.

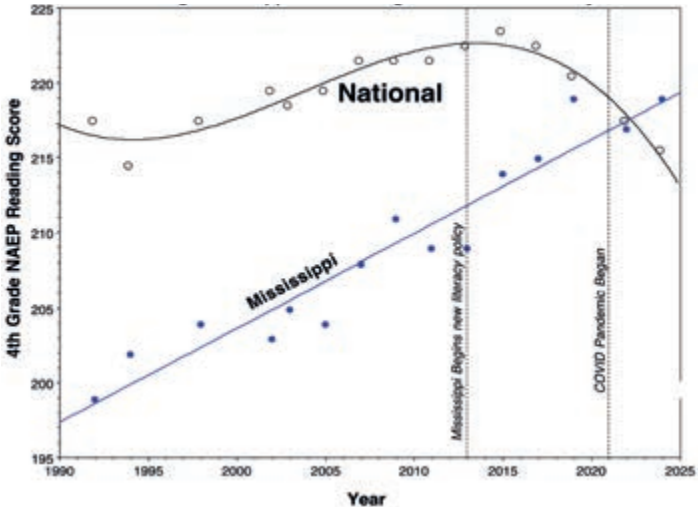

Figure 2: Mean fourth-grade NAEP literacy scores

how do we apportion that credit to the Act's two parts? Is it due to the changes in what was taking place in fourth-grade classrooms? Or who was allowed into those classrooms?

The most credible way to do this is with a formal experiment in which we form four groups by crossing the two factors – extra per-pupil expenditures, and promotion based on reading performance – yielding four experimental groups, and then assigning schools at random to one of those four groups. After the results from the fourth-grade NAEP test are known, we can easily calculate the effect of each of those two experimental factors. Unfortunately, in education, such experiments are rare (but when they are done – see the justly famous Tennessee class size experiment[7] – they offer clear, unambiguous answers to the questions of interest). In the current situation, the best we can do is to use the model given in Figure 1 to predict the gain in mean score from the retention rates and see how much of Mississippi's gains shown in Figure 2 are left unaccounted for. Were we to do this we would find that most of Mississippi's gains are due to the retention rate.

It is disappointing, but not surprising, that the lion's share of the effects of the "Mississippi miracle" are yet another case of gaming the system. There is no miracle to behold. There is nothing special in Mississippi's literacy reform model that should be replicated globally. It just emphasises the obvious advice that, if you want your students to get high scores, don't allow those students who are likely to get low scores to take the test. This message is not a secret.

Recently in Texas, the Houston school district again appeared in the news as having experienced another miracle. Since 2023, Houston had raised the percentage of high-school students meeting state standards in biology by 17 points and in Algebra I by 23 points. A miracle indeed. Or was it? Texas-based journalist Michael Hardy has exposed the man behind the curtain, a school district official. Houston pushed thousands of students into less rigorous biology and mathematics classes. This inflated test scores, as many of these likely lower-scoring students delayed taking the tests for at least a year. This sounds eerily similar to the 2013 third-grade Mississippi retention policy. Find a way to prevent the lowest test scorers from taking the exam and the average score will increase.

This is not the first or even the second time Texas has been caught gaming the system. In 2022, the city of Marlin, Texas had a parade to celebrate its best grade ever from the Texas Education Agency (TEA). The district had jumped from a 56 (F) in 2019 to an 86 (B) in 2022. Marlin was like most Texas public schools that increased their TEA grades in 2022. This occurred despite a decline in graduation rates and a decrease in ACT and NAEP scores. When school performance grades increase while other student performance measures decrease, something

It is disappointing, but not surprising, that the lion's share of the effects of the "Mississippi miracle" are yet another case of gaming the system. There is no miracle to behold

fishy is going on. How did this "miracle" happen? Robinson and Cole reported that Marlin's overall test score on the state-mandated State of Texas Assessment of Academic Readiness (STAAR) test was 56 in 2022.[8,9] It had been 50 in 2019. Not enough gain to celebrate with a parade. The overall TEA score of 86 was attributed simply to the boost generated by just 49 students who were given credit for taking a college prep course. That's it. Education reform achieved!

Back to the Mississippi Miracle and how Patrinos – head of the Department of Education Reform and 21st Century Endowed Chair in Education Policy at the University of Arkansas – and so many others could be so easily duped. There are three possible reasons. First, to his credit, Patrinos cited the 2024 study by Spencer[2] whose analysis concluded that the LBPA was the cause of the increase in fourth-grade reading and maths scores. The gold standard for measuring the effects of causes is to have a control group (e.g., students who continue with the old approach) and to randomly assign students to either treatment (the new instructional program) or control. The difference in the outcome between these two groups is the causal effect of the treatment relative to the control (see Chapter 11 in Wainer and Robinson[5]). Spencer did not have the data required for such an analysis. So, instead, he improvised by using some prior years' data as the control group, and instead of random assignment he used various bits of covariate information to equate this year's students with the previous years. Then he added the heroic assumption that the covariate information made two incomparable groups comparable (remember that in the current year the bottom of the class was truncated,

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 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). Dr. Robinson is the co-author (with Howard Wainer) of *Testing and the Paradoxes of Fairness* (Cambridge University Press, 2026).

making them very much unlike the prior years' scores – and no covariate adjustment was going to make them equal).

Second, besides weak empirical data, educational reformers like Patrinos should have given greater weight to the extant literature on the Mississippi Miracle. The miracle had already been convincingly debunked.[10] Fourth-grade gains had vanished by the time the students reached eighth grade.

Third, Patrinos, and others who have praised the Mississippi miracle, should know that extreme educational reform success stories are non-existent. History has shown us that a little bit of digging has, in the past, always revealed such claims of miracles to be false.[8,11] This does not mean that we should give up hope. Small successes are common in education. But dramatic huge successes should always alert us to scepticism.

Quo tendimus?

Where have our philosophical and statistical investigations of Mississippi's educational miracles brought us? And, more importantly, where are they leading? Let us conclude with two observations informed by these investigations.

The first is a possible future educational application. The development of the Gedanken Height School (GHS) was informed by Mississippi's educational program. GHS is a school whose sole purpose is to assist students in attaining their greatest stature. GHS guarantees that all of its graduates will be at least 6 feet tall. They propose to accomplish this with a programme of instruction focused on the value of additional height combined with required minimal test scores for promotion. To advance from ninth grade to tenth, a student must be at least 69 inches tall (shorter students will be retained); to progress from tenth to eleventh, a student must be at least 70 inches; and to move from eleventh grade to twelfth, they must have reached at least 71 inches. By the end of twelfth grade, there is a final evaluation to determine if a student is qualified to graduate, with the minimal passing score being 72 inches. GHS was a rousing success in that all graduates reached or surpassed the goals that were set for them four years earlier.

The second observation is a revised interpretation of education's past. More than a century ago, the enormous success of the industrial revolution naturally led educators to view the educational process through that lens. This vision imagined the school as a factory in which the students were the product. When the quality of the final product was judged flawed, pundits, politicians, and educators blamed the quality of the factory. This led to programmes to improve teachers, shrink class size, replace administrators, change educational philosophies, and so on. The educational landscape is littered with attempts to improve the factories' performance (e.g., the 1965 Elementary and Secondary Education Act, the 2015 Every Student Succeeds Act, the 2001 No Child Left Behind Act, and the American Community Survey launched in 2005) including, of course, various measures of students' performance (e.g., National Assessment of Educational Progress, Scholastic Assessment Tests (known as SATs), American College Testing (known as ACT), Massachusetts Comprehensive Assessment System (MCAS) and Iowa Tests of Basic Skills (ITBS)) as well as measures of associated teachers' performance (e.g., value-added models, Pre-Professional Skills Test, National Teacher Examinations, Praxis). The model of school as factory was swallowed whole and absorbed billions of dollars in pursuit of the success that always seemed to dance just out of reach. Over time, the public became accustomed to the failure of each new programme, so much so that when a story of success – even a modest one – emerged, it was characterised as a miracle. Witness the miracles claimed in Atlanta, New Orleans, Washington, DC, Texas, and, most recently, Mississippi. But, each time, a closer look at these extraordinary outcomes always affirmed Hume's observations about miracles. Yet there is a truth to be extracted from the factory model of schooling, for there is another way of improving factory output that could be considered. Experience has shown that factories can improve their products by upgrading the raw material on which they operate (the extent of improvement is made explicit in the equation we derived in Figure 1). The likelihood of its success can be seen in Mississippi's data – the schools perform best when students whose performances fall far below requirements are guided into different programmes that are judged to be better suited to their particular potential. But this is a tale for another day. ■

Since 2023, Houston had raised the percentage of high-school students meeting state standards in biology by 17 points and in Algebra I by 23 points. A miracle indeed. Or was it?

References

1. Hume, D. (2000) *An Enquiry Concerning Human Understanding* (Section X), Tom L. Beauchamp (ed.). Oxford: Oxford University Press. Originally published in 1748.

2. Spencer, N. (2024) Comprehensive early literacy policy and the "Mississippi Miracle". *Economics of Education Review*, **103**, 102598.

3. Patrinos, H. A. (2025) Mississippi's education miracle: A model for global literacy reform. *The Conversation*, 26 March. tinyurl.com/ack9kxfx

4. Robinson, D. H. and Bligh, R. A. (2019) Educational muckrakers, watchdogs, and whistleblowers. In P. Kendeou, D. H. Robinson and M. McCrudden (eds), *Misinformation and Fake News in Education* (pp. 123–131). Charlotte, NC: Information Age Publishing.

5. Wainer, H. and Robinson, D. H. (2026) *Testing and the Paradoxes of Fairness*. New York: Cambridge University Press.

6. Haney, W. (2000) The myth of the Texas miracle in education. *Education Policy Analysis Archives*, **8**, 41.

7. Mosteller, F. (1997) The Tennessee study of class size in the early school grades. *Bulletin of the American Academy of Arts and Sciences*, **50**(7), 14–25.

8. Robinson, D. H. and Cole, E. (2022) A "Texas miracle" brought us No Child Left Behind. Here's a new one. *Washington Post*. 30 October. tinyurl.com/5en6597k

9. Hardy, M. (2025) Houston ISD's Rising STAAR Scores Are Built On Holding Students Back. *Texas Monthly*, July. tinyurl.com/59weaa89

10. Hiltzik, M. (2023) How Mississippi gamed its national reading test scores to produce "miracle" gains. *Los Angeles Times*, 3 July. tinyurl.com/4n73hk4z

11. Heissel, J. A. and Ladd, H. F. (2018) School turnaround in North Carolina: A regression discontinuity analysis, *Economics of Education Review*, **62**, 302–320.