



Koretz, D. (2017). *The testing charade: Pretending to make schools better*. Chicago, IL: University of Chicago Press.

Pp. viii + 275

ISBN-13: 978-0-226-40871-2

**Reviewed by Audrey Amrein-Beardsley
Arizona State University
United States**

“Supporters of our current system will no doubt want to dismiss this book as yet another anti-testing or anti-accountability screed” (p. 8), just as they might want to dismiss this book review, written by someone who fully endorses and agrees with pretty much everything Koretz wrote in his “non-academic” (p. vii) book. But those quick to dismiss should realize that America’s test-based educational reform system is continuously being overdone with no evidence in support of its perpetual doing-ness. Approximately 90% of those who, along with Koretz and myself, define themselves as experts on educational tests and test-based policies would agree with this assertion, although nobody in the field “correctly predicted just how extreme the failures of test-based reform would be” (p. 243). This, in fact, is the key takeaway from this book, explained in more detail below, which should strike a chord with educational policymakers across the nation, should they not be quick to dismiss.

THE TESTING CHARADE



Pretending to Make Schools Better

Daniel Koretz

Albert Einstein, one of the most prominent and influential scientists of all time, left as part of his legacy a voluminous set of inspiring quotes¹ that also directly pertain to the contents of the book I review herein: *The Testing Charade: Pretending to Make Schools Better* (2017), authored by Dr. Daniel Koretz – Professor of Education and expert on educational tests and test-based initiatives and policies at Harvard University. Accordingly, I use a small handful of Einstein’s quotes to frame this book review, to help make evident what I consider to be some of the universal and scientific “truths” set forth by both Einstein and Koretz in these regards.

As per Einstein, the definition of insanity is to repeat the same behaviors over and over again in the hope that different results will materialize “the next time” after this or that is fiddled with or fine-tuned.² The theory of change at issue in Koretz’s book, which has been continuously reinvented over the past nearly 40 years as per American federal and state policies, suggests this: by holding districts, schools, teachers, and students accountable for meeting higher or “new and improved” standards, as measured by student performance on better or “new and improved” standardized tests, oft-with high-stakes or serious consequences attached to test output (e.g., graduation from high school, teacher and administrator pay, merit bonuses, school reconstitutions, school closures,

teacher tenure, teacher termination), administrators will supervise our public schools better, teachers will teach better, and students will take their learning more seriously. As a result, students will learn and achieve more, especially in America’s lowest performing schools, after which the nation will purportedly reclaim the global superiority³ perpetually desired.

Unfortunately, however, this theory of change continues to be revamped despite the paucity of empirical research evidence in support of its anticipated effects. Why? Simply put, it is overly “simplistic” (p. 195). Instead, it has “done a good deal of damage” (p. 195).

This is the primary point made evident throughout Koretz’s book,⁴ whereas after this 40 years of confidence in and experimentation with test-based educational reform policies as based on this theory of change, there is no doubt that “the whole idea of test-based accountability has failed...[and has]..become an end in itself, harming students and corrupting the very ideals of teaching” (book cover). Indeed, very few scientific studies have evidenced that this theory of change and its surrounding educational policies work. Rather, researchers have evidenced via countless scientific studies that this theory and its tangential policies, again repurposed over time, has caused unintended consequences instead, especially for the disadvantaged students that such initiatives were meant to help most.⁵ The

¹ See, for example: https://www.brainyquote.com/authors/albert_einstein

² There is some debate as to whether Einstein said this, or others including but not limited to Mark Twain and Benjamin Franklin did; however, the strongest evidence points to Einstein. See, for example, <https://www.quora.com/Did-Einstein-really-define-insanity-as-doing-the-same-thing-over-and-over-again-and-expecting-different-results>

³ For more information about how other countries use and rely upon tests and test-based educational policies and initiatives and reforms, or not, please

see Koretz’s Chapter 13: Doing Better (pp. 209-242).

⁴ Koretz defines this theory of change in different forms throughout the book writing one version of it, for example, as follows: “[T]he reforms seem to be simple common sense. Measure what is important, reward and punish people based on how much of it they produce, and they’ll produce more” (p. 32).

⁵ For more information about an also set of “praiseworthy” (p. 119) goals, for example, to reach targets or increase levels and percentages of “proficiency,” as typically arbitrarily defined, reduce the achievement gap (which has been

unintended consequences altogether outweigh the very few positive benefits⁶ realized post policy implementation.

On this point, Koretz “pulls no punches” (p. vii) as promised. He writes, for example, that “The evidence of these failures has been [continuously] accumulating,” yet the evidence is continuously and “routinely” being “ignored” (p. 7).⁷ Moreover, “the failures of the current system have festered as long as they have because many of the advocates of test-based accountability simply didn’t [and still don’t] want to face the evidence” (p. 245).

Related, and according to Einstein, “Not everything that can be counted counts, and not everything that counts can be counted.” This point pertains to a series of connected assumptions organized around what tests can and cannot do, also underscored throughout Koretz’s book. These assumptions relate to how those without knowledge, much less expertise in tests and their strengths and weaknesses (e.g., policymakers), naïvely trust and accept the output generated via tests as “truth.” Put differently, the assumptions gone wild here are that, because test-based measurements typically yield a mathematical and purportedly highly scientific value, a patent level of

certainty and exactness comes along with the numerical scores that result.

The fact of the matter remains, however, that “Even though measurement is scientifically based and rooted, [test-based] measurements function much more like language – they are essentially arbitrary designations that have no inherent value. Rather, their values are [socially] constructed given there is mutual agreement about how to use and interpret the numbers derived” (Amrein-Beardsley & Barnett, 2012, p. 2). Standardized achievement tests (and really all tests) are not all they are assumed to be, nor do they measure all that they are assumed to measure. Likewise, they do not often reveal what it is they are assumed to “objectively” or scientifically disclose. Indeed, we need to “pay attention to other important stuff” (p. 194)⁸ in that “even the best tests still leave a great deal unmeasured” (p. 195).

According to Koretz, test-based accountability has subsequently become “unmoored from clear thinking about what should be measured, how it should be measured, or how testing can fit into a rational plan for evaluating and improving our schools” (p. 5). Indeed, and noted earlier, this “overconfidence in standardized tests” (p. 195) has set America up for failure in that

narrowing between blacks and whites, although this started as early as the 1970s, and although the gap between the rich and poor has been consistently widening), or moreover “reduce the glaring inequities in the American education system” (p. 119), please see Koretz’s Chapter 8: Making Up Unrealistic Targets (pp. 119-136). For more about the achievement gap trends briefly noted above, please also see Koretz’s Chapter 11: Did Kids Learn More (pp. 175-192).

⁶ As per Koretz, “The best estimate is that test-based accountability may have produced modest gains in elementary-school mathematics [although the gains do not line up with the timing of the test-based reforms at the time; see, for example, p. 186] but no appreciable gains in either reading [across grades] or high school mathematics – even though

reading and mathematics have been its primary focus. These meager positive effects must be balanced against the many widespread and serious negative effects” (p. 6). See also Koretz’s more complete analysis of the intended effects of test-based policies and initiatives on student achievement in Koretz’s Chapter 11: Did Kids Learn More (pp. 175-192).

⁷ For more information about the points and assertions resident in this first section, please see Koretz’s Chapter 1: Beyond All Reason (p. 1-10) and Chapter 3: The Evolution of Test-Based “Reform” (pp. 21-36).

⁸ For more information about this other “stuff,” please see Koretz’s Chapter 12: Nine Principles for Doing Better (pp. 193-208).

“many of the people with their hands on the levers in education don’t understand what tests are and what they can and can’t do (p. 11). Rather, the “pervasive misunderstanding” of what tests can do and the widespread lack of understanding of what tests cannot do (e.g., count that which cannot be counted, capture and represent that which is actually *and* that which should be taught, including within and across other subject areas,⁹ and do all of this without error, bias, etc.¹⁰) are, consequently, what have quite literally kept us on this 40-year path of foolishness, tinkering toward some educational model of utopia (see also Tyack & Cuban, 1995).

Especially given the decades of human resource and financial investments invested into tests and the educational policies based on such tests, tests are just simply “so much less than those pushing for test-based accountability want them to be” (p. 15). Likewise, they certainly do not warrant some of their most “ludicrous uses” (p. 28), especially of late. Take, for example, the high-stakes evaluation and judgment of teachers’ allegedly causal impacts on students’ growth on such tests from one year to the next required of all states as per *Race to the Top* (RttT, 2011), and the recent *No Child Left Behind* (NCLB, 2001) waivers excusing states from not getting 100% of their students to 100% proficiency in reading and mathematics by 2014 (see also Philips, 2012).¹¹

Connected, and as per Einstein, “A man [sic] should look for what is, and not for what he thinks should be.” One should “Look deep into nature...[to]...understand

everything better.” In Koretz’s book, this ultimately relates to another useful section within this volume. This section has to do with Koretz’s summary of the past and current research on tests as framed using Campbell’s Law, throughout which Koretz calls for others to quite literally look deep into that which is perceived versus that which is real, especially in terms of the intended and unintended consequences of such test-based systems. This call is especially pertinent when high-stakes consequences are attached to, and then oft-distort, the “objective” test output blindly accepted as “truth.” Koretz accordingly argues that it is no question that “We have an *obligation* – [Koretz uses] that word deliberately – to examine the negative effects carefully and weigh them against the positive impacts to see whether the system is, on balance, succeeding” (p. 48).

According to Campbell’s Law (1976), “The more any quantitative [or qualitative] social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor” (p. 49). While in this case the quantitative indicators that educational policymakers endorse for inference- and decision-making purposes may perform well in low-stakes settings, as the stakes are raised, Campbell’s Law suggests that these indicators are increasingly susceptible to distortion.

When it comes to tests in education, not to mention many disciplines beyond education,¹² Campbell’s Law has been at play

⁹ Another Einstein quote is relevant here in terms of the blatant limitations of tests and that which they might represent: “Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.”

¹⁰ For more information about what tests can and cannot do, please see Koretz’s Chapter 2: *What Is a Test* (pp. 11-19).

¹¹ For more information about using teachers’ students’ test scores to evaluate teachers (e.g., using student growth models such as value-added models (VAMs)), please see Koretz’s Chapter 9: *Evaluating Teachers* (pp. 137-160).

¹² For more information about how Campbell’s Law comes into play across other disciplines, a when high-stakes consequences are attached to output elsewhere (e.g., emissions tests, health and

for years. Shepard (1990) and Haladyna, Nolen, and Haas (1991) who along with others (e.g., Linn and Madaus, as also noted by Koretz, pp. 39-40) formally introduced the phenomena surrounding test score inflation¹³ as also related to Campbell's Law and reviewed in this book.¹⁴ Instances whereby administrators and teachers have outright cheated¹⁵ and narrowed the curriculum to align with the curriculum tested; teachers have taught directly to the items tested (i.e., teaching-to-the-test, which is in fact bad practice despite some attempts to spin the contrary; see, for example, Popham, 2001); teachers have engaged in other problematic shortcuts and test preparation,¹⁶ test coaching, and test administration practices; and the like to increase test scores were all prophesied by Campbell's Law. Although others have effectively argued such points via noteworthy pieces elsewhere (see, for example, Haney, 2000; Jacob, 2005; Nichols & Berliner, 2007; Porter, 2015; Sidorkin, 2016), they are also effectively argued by Koretz throughout his book.

As it also was with states in the US using or still moving forward with the Common Core¹⁷, we cannot continue to simply “drop” ideas that might be reasonable and align with commonsense “into schools wholesale before we [gather] any evidence about impact; this has been true of almost the

entire edifice of test-based reform, time and again” (p. 164). Doing this is “analogous to a drug company saying that they have figured out, just based on their own beliefs and logic, which drugs will be effective and safe, so they can skip the time-consuming and expensive burden of actually gathering some evidence before selling” the drugs to consumers (pp. 175-176).

In sum, as per Koretz, as well as Einstein, we must stop the endless cycles of insanity when it comes to reforming America's schools using test-based reforms, policies, and initiatives. We need to be keeping in check what tests can and cannot do, while also acknowledging that “new and improved” tests will likely never improve upon America's capacities to do that which educational policymakers throughout the nation so desire them to do. All the while we should also be looking into what is, as compared to what should be, so as to understand everything better, for better and for worse, while also “accepting the need for human judgment” (p. 203) to help allow educators, as the professionals that they are,¹⁸ offset the unjustifiable and indefensible decisions too often based on such false perceptions of these test-based “truths.”

While the passage of the Every Student Succeeds Act (ESSA, 2016) represents

medicine, manufacturing), please see Koretz's Chapter 4: Campbell's Law (pp. 37-48).

¹³ For more information about test score inflation and pollution, please see Koretz's Chapter 5: Score Inflation (pp. 49-72).

¹⁴ For more information about how Campbell's Law comes into play, especially when high-stakes consequences are attached to test-based output, please see Koretz's Chapter 4: Campbell's Law (pp. 37-48).

¹⁵ For more information about cheating incidences and practices surrounding tests, please see Koretz's Chapter 6: Cheating (pp. 73-92).

¹⁶ For more information about how educators have literally “blur[red] – and in some cases entirely

obscure[d] – the distinction between test preparation and teaching” (p. 93), please see Koretz's Chapter 7: Test Prep (pp. 93-118).

¹⁷ For more information about the Common Core and its theoretical versus actual potentials to help the aforementioned theory of change work, please see Koretz's Chapter 10: Will the Common Core Fix This? (pp. 161-174).

¹⁸ For more information about how to do better by relying more on human judgment during the consumption and interpretation of test-based output, please see Koretz's Chapter 13: Doing Better (pp. 209-242).

the beginning of a shift to a more sensible and productive approach” (p. 246), Koretz warns us to not be fooled. ESSA’s basic test-based accountability and reform model, along with its reinventions based on a persistent and antiquated theory of change, returns to states throughout the US “just a fraction of the discretion” (p. 7) that they had prior in terms of their test-based educational reform policies.

ESSA certainly does not signal the end to this cycle of insanity. Rather, we must first start reforming that which we perpetually do (wrong) in this area of testing and test-based reform by “confronting honestly the failures that stare us in the face” (p. 9).¹⁹ Reading this book will certainly help us understand why reforming these failed reforms is so necessary.

References

- Amrein-Beardsley, A. & Barnett, J. H. (2012). Working with error and uncertainty to increase measurement validity. *Educational Assessment, Evaluation and Accountability*, 1-11. doi: 10.1007/s11092-012-9146-6
- Campbell, D. T. (1976). *Assessing the impact of planned social change*. Hanover, NH: The Public Affairs Center, Dartmouth College. Retrieved from http://portals.wi.wur.nl/files/docs/ppme/Assessing_impact_of_planned_social_change.pdf
- Every Student Succeeds Act (ESSA) of 2016, Pub. L. No. 114-95, § 129 Stat. 1802. (2016). Retrieved from <https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf>
- Haladyna, T. M., Nolen, N. S., & Haas, S. B. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7. doi:10.2307/1176395
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Analysis Policy Archives*, 8(41). doi:10.14507/epaa.v8n41.2000 Retrieved from <http://epaa.asu.edu/ojs/article/view/432>
- Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5-6), 761-796. doi:10.3386/w8968
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America’s schools*. Cambridge, MA: Harvard Education Press.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425. (2002). Retrieved from <http://www.ed.gov/legislation/ESEA02/>
- Philips, R. H. (2012, Sep. 10). No child left behind waiver sought by Alabama, six other states. *Press Register*. Retrieved from http://blog.al.com/live/2012/09/no_child_left_behind_waiver_so.html
- Popham, W. J. (2001). Teaching to the test? *Educational Leadership*, 58(6), 16-20.
- Porter, E. (2015). Grading teachers by the test. *The New York Times*. Retrieved from <http://www.nytimes.com/2015/03/25/business/economy/grading-teachers-by-the-test.html>
- Race to the Top Act of 2011, S. 844—112th Congress. (2011). Retrieved from <http://www.govtrack.us/congress/bills/112/s844>

¹⁹ For more information about how American might improve upon its testing and test-based accountability systems, given the likelihood that despite the evidence they will likely continue to persist (see, for example, the note about ESSA

(2016) prior), please see Koretz’s Chapter 12: Nine Principles for Doing Better (pp. 193-208) and Koretz’s Chapter 13: Doing Better (p. 209-242). See also Koretz’s Chapter 14: Wrapping Up (pp. 243-248).

- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9, 15-22. doi:10.1111/j.1745-3992.1990.tb00374.x
- Sidorkin, A. M. (2016). Campbell's Law and the ethics of immensurability. *Studies in Philosophy and Education*, 35(4), 321-332. doi:10.1007/s11217-015-9482-3
- Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

About the Reviewer

Audrey Amrein-Beardsley is currently a Professor in Mary Lou Fulton Teachers College. Her research interests include educational policy, educational tests and measurements, educational policies as based on tests, and research and evaluation methods. Her research also focuses on the use of value-added models (VAMs) in and across states before and since the passage of the Every Student Succeeds Act (ESSA).




education review // reseñas educativas

a multi-lingual journal of book reviews

editors: **gustavo e. fischman /
melissa cast-brede / gene v glass**

Supported by the Mary Lou Fulton Teachers College, Arizona State University

 *Education Review/Reseñas Educativas/Resenhas Educativas* is supported by the edXchange initiative's Scholarly Communications Group at the Mary Lou Fulton Teachers College, 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, and distribute this article, as long as the work is attributed to the author(s) and ***Education Review***, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or ***Education Review***. ***Education Review*** is published by the Scholarly Communications Group of the Mary Lou Fulton Teachers College, Arizona State University.

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*.

Connect with *Education Review* on Facebook (<https://www.facebook.com/pages/Education-Review/178358222192644>) and on Twitter @EducReview