



The disjuncture between raw scores and pass rates in New York State public schools: Turning success into failure



William Mangino^{*}, Marc Silver, Jonathan Cavalieri¹

Hofstra University, USA

ARTICLE INFO

Article history:

Received 14 August 2014

Received in revised form 12 February 2015

Accepted 17 March 2015

Available online 4 April 2015

Keywords:

Test score gap

Cut scores

Failing schools

High stakes testing

Proficiency

Standards based reform

ABSTRACT

This paper demonstrates that ‘failure’ is not a direct reflection of student knowledge. Using five years of New York State school-level data, we compare *passing rates* to *raw-scores*. We find, first, that when ‘cut scores’ are raised, more students fail even if raw scores are increasing. Second, increasing cut scores disproportionately fails more poor students than non-poor students, despite that poor students have the fastest rates of raw score improvement. Third, raised cut scores transform the *smallest* raw score gaps between high- and low-poverty schools into the *largest* passing gaps. Thus, while students in poor schools know more than they did previously, and although they have learned at superior rates, they are recast as the biggest ‘failures’ they have ever been.

© 2015 Elsevier Ltd. All rights reserved.

Introduction: the reification of failure

At least since *A Nation at Risk*, and coming to full fruition in No Child Left Behind (NCLB), Race to the Top, and now the Common Core, “standards based reform” is the dominant ethos of American educational innovation (Baker & LeTendre, 2005). The central premise of standards based reform is that by raising academic standards, and then holding students, teachers, administrators, schools, districts, and states accountable for student performance, students will learn more material, and will learn it better (for a reviews, see Hamilton, Stecher, & Kun Yuan, 2012; McLaughlin & Shepard, 1995). Thus standards based reform, the argument goes, will make individuals, localities, states, nations, and regions more competitive in the modern global economy (Grubb & Lazerson, 2004). In the words of the USA’s 43rd President, doing otherwise amounts to the “soft bigotry of low expectations” (Bush, 2000).

For better and worse, standards based reform and accountability have gone hand-in-hand, and in this context, accountability means high-stakes standardized testing (Carnoy, Elmore, & Siskin, 2003), where district- and school-level, and increasingly individual

level, data are released to the public, describing who is ‘proficient’ and who ‘fails’. Data at the aggregated levels commonly identify whole schools and districts as failures, and such breakdowns are reported via school report cards and paraded in local newspapers and news programs (e.g., *New York Times*, 2010, 2007; *Newsday*, 2013a, 2013b). It becomes public knowledge, quickly tuning into ‘common sense’ where the ‘good schools’ are, and where the ‘bad schools’ are.

Until recently with many states joining the Common Core movement, there were wildly differing definitions of ‘proficient’—the mark of passing—as NCLB left such definitions to each state to decide (Cronin, Dahlin, Adkins, & Kingsbury, 2007). In order to meet federal mandates and not lose federal funds, some states simply made accountability tests easier to pass (Carey, 2006)—the exact opposite of what standards based reform was supposed to do in the first place. Since states established their own varying definitions of proficiency, some scholars mapped them on to a common scale and found that proficient in a given subject can be as low as the 6th percentile to as high as the 77th percentile when the National Assessment of Educational Progress (NAEP)² is used as the reference (Finn & Petrilli, 2008; see also Bandeira de Mello,

^{*} Corresponding author at: Hofstra University, Department of Sociology, 205 Davison Hall, Hempstead, NY 11549, USA. Tel.: +1 516 463 5640.

E-mail address: william.mangino@hofstra.edu (W. Mangino).

¹ Formerly at affiliation ‘a’. Now at: Vincentian Institute for Social Action at St. John’s University, USA.

² The NAEP, administered by the National Center for Education Statistics, is the only test that is given to a representative sample of students from all states. As such, it provides statistically valid comparisons of test scores across states, and it can be used as a common scale to compare state-defined pass rates.

Blankenship, & McLaughlin, 2009). While there is evidence that higher standards can improve pass rates for “struggling students” (Clark & Cookson, 2012) and that private school students learn more than public school students because there is less choice to opt out of difficult courses (Carbonaro & Covay, 2010), an important finding of the research on proficiency is that the “rigor of the state standards is not consistently associated with higher performance. . . . Most of the variation. . . in the percentage of students scoring proficient or above on state tests can be explained by the variation in the level of difficulty of state standards for proficient performance. States with higher standards. . . had fewer students scoring proficient on state tests” (Bandeira de Mello et al., 2009, p. vi). More simply, when tests are harder, fewer students are proficient, and when tests are easier, proficiency increases. By implication, then, it is possible that observed differences in aggregated pass rates might not reflect differences in student knowledge; rather they might be the product of differing *cut scores*.

Just prior to adopting the Common Core, New York State (NYS) implemented new academic standards. A change was made in 2010, when the State Education Department changed the scores that define passing on the state’s yearly standardized tests (NYSED, 2010a, 2010b). These changes, described in detail below, present a unique research opportunity insofar as at this stage of reform the content of tests was not changed, just the passing score was revised. That passing score, called the cut score, is a division that cuts the test-score distribution in two; above the cut are passers and below the cut are failures. While the division between pass and fail seems obvious, it is actually quite problematic. Sociologist Pierre Bourdieu has called attention to this line of pass and fail and the very real distinctions that it makes. Bourdieu states,

One has only to think of the. . . competitive recruitment examination which, out of the continuum of infinitesimal differences between performances, produces sharp, absolute, lasting differences, such as that which separates the last successful candidate from the first unsuccessful one, and institutes an essential difference [of] the officially recognized, guaranteed competence. (Bourdieu, 1986, p. 248, italics added)

These are powerful words, and worth reiterating. Think about the comparison between someone who gets the lowest passing score on an exam, say 60% correct, versus the highest failing score, say 59% correct. While there is no pragmatic difference in knowledge represented by these two scores—they are certainly statistically same, within a normal margin of error—real and important differences are nonetheless *instituted* (hence Bourdieu calls such distinctions “institutionalized capital”). Things like who gets retained in a grade and who gets promoted to the next; who gets what remedial services and who does not; who watches their friends go on, while they do not; whose self-esteem is bolstered and whose decays; who is ‘proficient’ and who is a ‘failure’—all this because of the distinction derived from a cut score.

The identification and implementation of a cut score that divides passing from failing is a form of the *reification* fallacy. We do not use the term, reification, in a strict Marxian way (for a review and reframing, see Honneth, 2008), although our argument could be so framed. Rather, we are closer to a standard dictionary definition of “regard[ing] (something abstract) as a material or concrete thing” (Merriam-Webster, 1987, p. 993). Commonly, ‘failure’ is looked at as a real and objective indicator of academic performance, but it is not; failure is a social construction that is both the product of history and a bureaucratic/political artifact. We use the term, reify, in a statistical sense, like a ‘Type I’ error, but statistical significance is not the concern. It is the idea that ‘what gets measured becomes real’; it is to ‘thingify’. One of the best usages of the term in this sense is Stephen Jay Gould’s (1996, especially chapter 6, which is subtitled “Factor Analysis and the

Reification of Intelligence”) critique of the Intelligence Quotient (IQ). In Gould’s view, just because psychometrics can produce a measure called “IQ”, it does not mean that IQ—a singular, innate, unchanging number that represents general ability—is a valid or real thing. But nonetheless, IQ is widely used; it takes on a perceptual and popular ‘realness’ and authority that it in fact does not have. The same use of the word shows up in medical literature. It seems physicians place more credence in numbers than they do in narratives (Sorensen, 2003); and while not using the word, Best has written books on how people “accept even the most implausible [statistics] without question. . . . [Such statistics] can distort our understanding of our world; and they can lead us to make poor policy choices” (Best, 2001, pp. 4–5). This is what we mean by the reification of failure. It is an instituted distinction that solidifies, simplifies, and makes concrete varying educational performances that are in fact much more nuanced. Like many works in the Bourdieu tradition (e.g., Rafferty & Hout, 1993; Soares, 2007, 2012) this paper shows how seemingly neutral academic standards, like cut scores, are treated by authorities and the public alike as meritocratic and real, while they actually can serve to maintain and magnify social inequalities.³

To show this dynamic in NYS, the cut score must be considered in reference to the *raw score*. Simply, the raw score is a student’s actual score on an examination. This is what we normally think of when we discuss a student’s percent correct on an exam (although raw scoring does not have to be expressed in percentages). The raw score does not denote passing and failing, per se; it just indicates ‘how much knowledge’ a student demonstrates (leaving aside that it also measures knowing ‘how to take a test’). Raw scores are what creates Bourdieu’s “continuum of infinitesimal differences”.

A standardized test is, by design, comparable from one administration to the next. So, on a given, annual standardized test, raw scores provide year-by-year measures of how much students in a given grade in a given year actually know in a given subject. While raw scores are comparable across years (assuming proper ‘standardization’, i.e., the content of the test has not changed. See Rosner, 2012 on how standardization is achieved and the pitfalls thereof), the percentage of students who pass an exam is *not* necessarily comparable across years, because educational administrators can change the cut score from year to year, thereby changing passing rates without any commensurate changes in knowledge.

It is this distinction between knowledge (raw scores) and how much knowledge is deemed proficient (cut scores) that drives the analysis in this paper. We show that implementing higher academic standards by raising cut scores, while applied equally to everyone, disproportionately fails more at-risk students and schools. The irony is that even as raw scores are rising—showing that improvements in knowledge have been attained—and the gap between at-risk and not-at-risk tests-takers has been shrinking in terms of knowledge, the implementation of higher standards via raising cut scores creates disproportionately more failures among poor and non-white students, turning them into failures, despite their increasing levels of knowledge.

To show these mechanisms in New York State, our statistics are simple, but insightful. We will compare ‘high’ poverty schools to ‘low’ poverty schools, demonstrating that high poverty schools are set back more by the increase in cut scores than are average schools or low poverty schools. We will show that despite continual narrowing of the gap on raw scores—that is, high poverty schools are improving on tests more rapidly than are low poverty schools—raising cut scores artificially increases the gap independently of the amount of knowledge actually demonstrated by students. The

³ This sentence’s phrasing in part comes from a blind reviewer for *Studies in Educational Evaluation*.

Download English Version:

<https://daneshyari.com/en/article/372613>

Download Persian Version:

<https://daneshyari.com/article/372613>

[Daneshyari.com](https://daneshyari.com)