

# The Roots of Grades-and-Tests

Introduction to *De-Testing and De-Grading Schools: Authentic Alternatives to Accountability and Standardization*, edited by Joe Bower and P.L. Thomas  
(Peter Lang Publishing, 2013)

---

## The Roots of Grades-and-Tests

By Alfie Kohn

Most of the contributions to this book focus on problems with either grades or tests. In an article about college admissions published more than a decade ago, however, I suggested that we might as well talk about “grades-and-tests” (G&T) as a single hyphenated entity.[1] There are certainly differences between the two components, but the most striking research finding on the subject is that students’ G&T primarily predicts their future G&T – and little else. It doesn’t tell us much at all about their future creativity, curiosity, happiness, career success, or anything else of consequence.

In fact, the case for the fundamental similarity of grades and tests runs deeper than their limited predictive power. Both are “by their nature reductive,” as P. L. Thomas, the editor of this anthology, observes in his chapter. I would add that both emerge from – and, in turn, contribute to – our predilection for three things: quantifying, controlling, and competing. All of these are defining characteristics of our educational system but also of our culture more generally.

To quantify is to talk about something in numerical terms.

That's not a problem when a question lends itself to counting ("How large are elementary school classrooms as opposed to high school classrooms?") but becomes more troubling in the case of other inquiries ("How do we know if that teacher is any good?") Just over a hundred years ago, Edmond G. A. Holmes, the chief inspector of elementary schools for Great Britain, remarked, "As we tend to value the results of education for their measurableness, so we tend to undervalue and at last ignore those results which are too intrinsically valuable to be measured." [2]

Tests, or at least those that yield a score, are, like grades, based on the premise that learning can and should be quantified. Indeed, the pervasiveness of G&T suggests that the (reasonable) question "How should we assess...?" has morphed into the (more problematic) question "How should we measure...?" – as if assessment without numbers was either (a) so obviously inferior as to be undeserving of discussion or (b) simply impossible, because to assess *means* to measure. [3]

In his book *Trust in Numbers*, historian Theodore Porter points out that quantification has long exerted a particular attraction for Americans. "The systematic use of IQ tests to classify students, opinion polls to quantify the public mood,...even cost-benefit analyses to assess public works – all in the name of impersonal objectivity – are distinctive products of... American culture." [4] I don't know whether this is more true in education than it is in other fields, but it seems particularly disturbing to assume that the process by which children come to make sense of ideas is always something we can count. And that assumption reveals itself not only through the ubiquity of G&T but also through more recent (and equally reductive) developments such as rubrics, which have the effect of smuggling in standardization through the back door. [5]

The enterprise of assessing and evaluating requires teachers to do two things: collect information about how students are

doing and then share that information with the students and/or their parents. But tests aren't necessary to do the first and grades aren't necessary to do the second. A teacher who is paying attention – listening to students' conversations, following their projects, reading their writing – will never need to administer a test. (Of course, this assumes that students have a chance to converse, design projects, and write. If they're forced to spend their time listening and filling out worksheets, well, then there's not much authentic learning to *be* assessed.) In fact, that attentive teacher will acquire a broader and deeper understanding of how her students are faring, and which of them need help with what, than she could with a test.

Tests are not only unnecessary but unhelpful because they mostly tell us how many forgettable facts have been crammed into short-term memory, and how skilled students have become in the specialized art of test-taking. Steven Wolk, an Illinois teacher, put it this way:

In the real world of learning, tests and reports and worksheets aren't the most meaningful way to understand a person's growth, they're just *convenient* ways in a system of schooling that's based on mass production....I assess my students by looking at their work, by talking with them, by making informal observations along the way. I don't need any means of appraisal outside of my own observations and the student's work, which is demonstration enough of their thinking, their growth, their knowledge, and their attitudes over time.[6]

Once the teacher has figured out the extent to which students' thinking is becoming more sophisticated and where gaps still exist, there's obviously no need to reduce the conclusion to a summary letter (B) or a number (84) or a label that functions just like a letter or number but allows us to pretend we're doing something different ("exceeds expectations"). Instead, a qualitative description or evaluation can be offered in

narrative form – or, better yet, as part of a dialogue during a meeting with students or parents.

Why is G&T still so common if it's unnecessary and, as many of the chapters that follow in this book argue, downright harmful? Possible answers include: tradition, the appeal of quantification (with its siren call of objectivity), a lack of familiarity with alternatives, and, as Wolk points out, simple convenience. But here's another explanation: Unlike more authentic ways of determining and then describing students' progress, G&T appeals to those who seek control. If I don't know how to work with my students to create a classroom and a curriculum that will pique their intellectual curiosity and persuade them to participate, I can simply coerce them into doing whatever I say – show up on time, sit down, and be quiet; write down what I say; read these pages or complete these exercises (at a pace I impose); do even more schoolwork at home – by warning them that noncompliance will result in their faring poorly on a test, which, in turn, will bring down their grade.

Extrinsic inducements, of which G&T is the classic example in a school setting, are devices whereby those with more power induce those with less to do something. G&T isn't needed for assessment, but it is very nearly indispensable for compelling students to do what they (understandably) may have very little interest in doing. The same is true of standardized tests as a matter of public policy, particularly when rewards or punishments hinge on the results. This is how federal officials make state officials race to what they define as the top, how state officials make district administrators adopt a set of prescriptive curriculum standards, how administrators deprofessionalize teachers by compelling them to follow scripted lessons, and so on. (For readers who are already familiar with how high-stakes testing serves as a mechanism of control, what may be the new insight here is that the same is true of teacher-designed tests and quizzes, which are

instruments by which teachers treat their students much as they complain about being treated themselves.)

As the engine of both school “reform” at the macro level and in-class assessment at the micro level, then, G&T creates spurious precision, flattening education into something that can be measured, and forces people to participate whether they like it or not. But it also has a third effect, which is to foster competition. Educational and psychological tests were invented to sort people – not just to rate but to rank. The original imperative wasn’t to learn about test-takers in order to help them, but to determine who was better than whom and, practically speaking, which of them to select and which to leave behind.

Despite this history, it is possible to test in such a way that the results will not be used to pit students against one another for recognition or rewards – although testing remains problematic for other reasons. Similarly, grading needn’t be done on a curve; the system can be set up so all students, at least in theory, may earn the top grade. (Grades would still function as extrinsic motivators but at least there wouldn’t be an artificial scarcity of A’s.) Yet in practice G&T never seems to be too far removed from competition: Quantified results create an irresistible temptation to compare students. Even schools that prohibit teachers from grading on a curve may use grades to compute class rank, and the students themselves may feel compelled to keep asking one another, “Wad-ja-get?”

It’s not a coincidence that defenders of G&T point to our competitive culture (recast as “the real world”) in order to justify the practice. At the same time, those who are troubled by the effects of competition tend to be critical of G&T as well, and vice versa. Specifically, teachers who are committed to cooperative learning (as well as to democratic classrooms and the kind of thinking that can’t be reduced to numbers) are also, in my experience, apt to steer clear of G&T

whenever possible.

The distinguishing feature of that opposition is that G&T, and the underlying adherence to quantification, control, and competition, is understood as a problem in itself. We have to look beyond real but marginal objections to the way G&T has been implemented. The problem with testing isn't limited to what's *on* the test (or, even less important, whether the results are released in time to "do any good"). The problem with grading isn't limited to how many students get A's, or what role homework or class participation plays in determining the final grade, or whether it's possible to retake a test, or whether marks are posted on-line. Nor will replacing norm-referenced with criterion-referenced tests, or letter grades with rubrics, do the trick. The problem runs deeper, so our willingness to question and confront the status quo must follow suit.

## NOTES

1. Alfie Kohn, "Two Cheers for an End to the SAT," *Chronicle of Higher Education*, March 9, 2001.
2. Edmond G. A. Holmes, *What Is and What Might Be?* (London: Constable, 1911), cited in George Madaus and Marguerite Clarke, "The Adverse Impact of High-Stakes Testing on Minority Students," in *Raising Standards or Raising Barriers?* (New York: Century Foundation Press, 2001), p. 93.
3. For more on this topic, see my short essay "Schooling Beyond Measure," *Education Week*, September 19, 2012.
4. Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, NJ: Princeton University Press, 1995), p. 147.
5. See, for example, Maja Wilson, *Rethinking Rubrics in*

*Writing Assessment* (Portsmouth, NH: Heinemann, 2006).

6. Steven Wolk, *A Democratic Classroom* (Portsmouth, NH: Heinemann, 1998), pp. 111-12.

Copyright © 2013 by Alfie Kohn. This article may be downloaded, reproduced, and distributed without permission as long as each copy includes this notice along with citation information (i.e., name of the periodical in which it originally appeared, date of publication, and author's name). Permission must be obtained in order to reprint this article in a published work or in order to offer it for sale in any form. We can be reached through the [Contact Us](#) page.