

# THE K–12 MATH TEST CONUNDRUM

FRANK QUINN

## 1. INTRODUCTION

This is a story with disasters, well-intentioned efforts gone wrong, some villains, and no quick fix. There are small constructive steps, however, mostly concerned with having higher-quality tests.

## 2. TESTS

High-stakes K–12 math tests are necessary. Math skills of high-school graduates are falling even as the world becomes more technical. Ranking on international tests is a national embarrassment. Many colleges' math offerings are dominated by remedial courses. The US imports high-quality education to fill its graduate programs and provide new faculty and a high-tech workforce. None of this will change without high-stakes tests to provide discipline and accountability.

High-stakes K–12 tests are also an abomination. They are high-stakes for schools as well as students so administrators apply pressure to teachers to raise scores. Teachers do their best to ensure students are tense and motivated. As a matter of survival teachers teach to the test, so things skipped on the test are left out of the class. And if “understanding” doesn't seem to help scores then tricks and mechanical drill are the order of the day.

The tests themselves are antiques: one-shot tests with multiple choice questions and op-scan answer sheets. Questions are mediocre, often ambiguous, and sometimes wrong. Question choices skew the curriculum since teachers teach to tests. Even the answer format has consequences. Questions are often numerical, in part to avoid giving clues in answer choices:  $\pi r^2$  is obviously the area of a circle, but 16.6 is not obviously the area of a circle of radius 2.3. This, and calculators, have shifted focus to numerical work and led to a significant decline in abstract and symbolic thinking.

The antique-test problem, at least, has an explanation and some prospect of relief. Since each edition of a traditional test is essentially made from scratch it is necessary to weight scores to keep outcomes consistent. Determining this weighting is by far the greatest cost of test development (sometimes hundreds of dollars per question), and the testing industry has an enormous investment in the necessary expertise. However this is irrelevant to modern multi-try computer-based tests repeatedly using a single system. Each student gets a slightly different test, and if these vary a bit in difficulty (or the student has a bad day) the remedy is to retake it. It seems unlikely the traditional industry will make this transition since it would mean discarding their major asset. However a new industry is developing and soon

the only barrier will be the courage needed for a department of education to break with their traditional partners.

### 3. STANDARDS

Each US state has a standards document for mathematics education. In principle this organizes teaching, texts and tests, and ought to ease problems with tests by providing an environment designed to prepare students for them. The dysfunctional current situation would thus seem to be due to dysfunctional standards documents, and consequently much of the debate (including the “math wars” of California and elsewhere) is focused on them. There is more to it than that, but these documents certainly do leave a lot to be desired.

Most states’ standards documents have not been reviewed by an academic mathematician, and are rife with inaccuracies. Many are so ambitious they cannot have been reviewed for realism by classroom teachers. A few set goals so low that college preparation must not be a concern. It is common to describe material in “bands” of several grades. It is then consistent with the standard to promote students who have learned very little in the first year of the band, guaranteeing a lot of repetition in the last year. Sometimes material is described in topic-centered “threads” with no due date at all.

Most standards are largely concerned with untested material. They specify process: students should learn about addition by playing with rulers. They specify exposure not expected to result in testable skills. The word “understand” for instance is clearly distinguished from “able to work problems with”. This empowers teachers to promote students who “understand” material if they can’t work problems. There is a natural sense of betrayal when this presumably higher-level knowledge is not recognized by high-stakes tests.

### 4. WHAT WE CAN DO

The current state of US math education is a national emergency, and our location in the educational system gives academic mathematicians a crucial role. It is a delicate role, however, since there is no agreement on the nature or even the existence of the emergency. There are politicians who feel that their election to high office after failing algebra demonstrates that math is not necessary for success. The K–12 community for decades has been shifting to qualitative and imaginative understanding, and a good deal of the estrangement from the academic community results from our reactionary attachment to testable skills and precision. Some suggestions:

- If you have an opportunity to participate in development of state standards, take it. But be prepared to yield on many issues. Remember this is an incremental process, and if our viewpoints are correct then the discipline imposed by high-stakes tests give us a long-term advantage.
- If you have an opportunity to discuss K–12 pedagogy, pass. There are exceptions to this of course, but at the moment this is a nearly religious issue and alienation is more likely than progress. Again the discipline of high-stakes tests should help as the issue becomes more a struggle against a common enemy and less a difference of opinion or world view.
- Think about tests. Tests will be driving standards and curriculum development in the near future since they provide concrete objectives and

quantitative measures of success. But bad tests drive the process in unfortunate directions and the current test-development process is a serious weak point. Large collections of mathematically “wholesome” sample problems would be a great resource. There are particular needs for good word problems with symbolic (non-numerical) outcomes.