DYSFUNCTIONAL STANDARDS DOCUMENTS IN MATHEMATICS EDUCATION

FRANK QUINN

Abstract. Standards documents attract a great deal of attention, and reasonably so: they should provide structure and common reference points for teachers, administrators, curriculum developers, textbook writers, test developers, etc. Unfortunately current documents do a poor job with all this and it seems unlikely they will improve.

Contents

1. Introduction

This essay grew out of a meeting held in Park City in July 2004 to review K–12 Mathematics Standards Documents from the 50 US states. The objective was to look for commonalities and in particular determine the extent to which the 1989 NCTM publication *Curriculum and Evaluation Standards for School Mathematics* had acted as a template for the development of such commonalities. In fact this publication has been very influential, but it has been interpreted in so many ways—particularly in the upper grades—that the commonality seems to be more of language than content. Further there seemed to be little reason to think this influence has made these documents more effective.

Here we back up a bit and ask: what is the point of a standards document? Who is supposed to read it, and what do they need from it? In §2 we list some jobs to

Date: December 2004, revised November 2008.

be done. These jobs fall to the standards document by default—nothing else can do them—so standards writers cannot decide whether or not to take on these jobs, they can only decide how well to do them.

§3 concerns standard high–stakes tests. These seem to be necessary as the "enforcement arm" of the standards process[1], but current tests are often counterproductive and linkage to standards is poor.

Even this brief discussion makes clear that genuinely effective documents will be very difficult to develop and are beyond the scope of current procedures. We may have to seek other ways to improve our educational system.

2. Roles for Standards Documents

Standards documents provide organizational frameworks and common reference points for teachers, administrators, curriculum developers, textbook writers, test developers, and indeed anyone interested in the system. To be effective they must be interpreted unambiguously and consistently by all parties.

2.1. **Tests.**

The most problematic relationship is between standards and tests. In principle the high-stakes system-wide tests coming into wide use should reflect standards. In practice standards are vague and often unrealistic statements of goals that give little hint how they should translate to a test.

For example some state standards include enough probability and statistics to justify the use of a college final exam as the state test. This is clearly unrealistic, and in fact very little of the material appears on actual tests. Teachers obviously find old tests much better guides to the outcomes expected. Subsequent tests are expected to be consistent with the old ones, and to accomplish this the test designers refer not to the standards, but to the old tests. As a result old tests become de facto standards. To the extent the official standard differs from tests it becomes irrelevant.

There are many well-known disadvantages to teaching to a test, and some are discussed in §3. But with high-stakes testing this will happen unless a teacher can use the official standard to anticipate tests in detail: what will *not* be covered as well as what might. At the very least this would require a large number of sample problems and careful attention to what can be realistically accomplished in a typical classroom.

2.2. **Textbooks.**

One of the main influences of standards documents is guiding the selection of textbooks. But conformity to standards is difficult to determine in the best of circumstances and nearly impossible when standards are unclear. To deal with this some publishers and state departments of education have developed a bizarre convention: the publisher prints at the top of each page the standard ostensibly addressed by the material on the page. The department of education then checks to be sure each standard appears at the top of at least one page. This can hardly be thought of as quality control.

Linking texts and standards can also enforce a disconnect between texts and teaching: teachers take old tests as de facto standards, tests differ from official standards, and texts conform to official standards. In any case the standard does not serve as a useful common reference point.

[1] See The K–12 Math Test Conundrum for a brief discussion.

Matching texts to standards would be possible if standards are detailed and stable, and worthwhile if they also correspond to classroom practice.

2.3. **Coordination.** A common observation in comparisons of educational systems is that US programs have far more repetition. Some of this is by design but some may be a consequence of the standards system.

Frequently material is described in multi-year "bands" or "threads" rather than for years or smaller increments. This means teachers cannot count on material having been mastered until the end of the band. Or put another way, it is consistent with the standard for teachers early in the band to pass students who have not assimilated the material, essentially guaranteeing that it will have to be repeated before it can be used later.

To avoid repetition a standard must not only specify the learning goals in a class, but also enforce discipline in getting it done by specifying that the material *not* be repeated in non-remedial classes later in the curriculum. In the absence of standard curricula, standards documents are the only way this sort of coordination can be accomplished.

Standards can coordinate content as well as timing. Material learned in early grades is needed in later grades. Material learned in later grades is needed in college or the workplace. Unfortunately it is common to find that teaching methods or simplified problem sets focused on a particular level do not effectively support the needs of later levels. Standards documents are the ideal place to address this.

For instance when specifying that multiplication of multi-digit numbers should take place at one level, the standards might also recommend that this be done in a way that will support multiplication of polynomials at a later level. Or when students first learn to factor polynomials they usually see many with integer roots because these are easy to do. Some students get the unfortunate impression that quadratics usually have integer roots. A standard could have a warning about this and require a significant number of problems with irrational roots.

For standards to be successful in coordinating a program they must be detailed, explicit, and stable.

2.4. **Process and Outcomes.** The discussion above suggests that to be effective a standard should describe testable outcomes in considerable detail. In some cases testable outcomes and non-testable supporting activities occur in different courses or grade levels and organizing this may be part of the job of a standards document. However for the most part prescribing non-testable activities is likely to be counterproductive:

- Most teachers feel that figuring out how to meet testable goals via non-testable activities is a pedagogical or curricular issue properly the domain of teachers.
- On a practical level anything not explicitly labeled "not tested" is a liability for teachers: there is always a risk that some test writer will figure out how to test it in an unexpected way.
- Non-testable goals may act as loopholes through which students receive credit and advance without acquiring skills. Lower skill expectations are certainly appropriate for some students but this should be explicit and managed rather than hidden in loopholes.

As an illustration of the last point, it sounds right to say "understanding is more important than rote mechanical skills." However there are several ways to interpret this. College teachers would take "understanding" to include effective skills, so an inability to work problems implies a lack of understanding. In contrast standards documents almost universally use "understand" to mean "exposed to but not expected to work problems with". "Know" is frequently used the same way though occasionally it means "able to reproduce" (as in "the student will know the formula for the area of a rectangle") or "able to identify among three alternatives" (on a multiple–choice test). In this interpretation "know" and "understand" are not linked to testable skills[2].

Students, teachers and parents may reasonably infer that "understanding" is a separate—and possibly superior—pathway to success, distinct from mere skill acquisition. They feel cheated that high–stakes tests and college teachers do not reward such understanding. However when the rubber hits the road in later courses or real life, the skills needed are the ones that can be tested. Students promoted on the basis of nonfunctional understanding are at a disadvantage.

These problems can be avoided if standards documents focus on testable outcomes. Sometimes more careful use of language may help. However this not the whole solution since legalistic precision often leads to legalistic obscurity, and is more useful for fixing blame than for preventing problems. A better approach would be to illustrate every testable expectation with a representative sample problem, and explicitly link untestable activities to testable ones later in the curriculum.

2.5. **Mathematical Structure.** The points above are not subject-specific and may apply to other problem-oriented subjects. Mathematics does have some subject-specific features: first it is *cumulative* in that *essentially all* knowledge and skills learned at one level will be needed at later levels. Second it has a lot of *abstract logical structure*. Teaching mathematics, and therefore any document that structures the teaching of mathematics, should be consistent with these features.

We expand on the role of structure. In practice most math problems are routine applications of mechanical skills. These skills really are needed in later work, and few K–12 students are able to effectively learn abstract structure, so skills are an appropriate focus. However it is abstract structure that makes mechanical routines work, and the better they reflect the structure the better they work. Further most students internalize abstract structure if it is clearly displayed in routine work. This internalization makes it easier to progress to deeper work based on similar ideas, and eventually to the ideas themselves.

As an example, the arabic digit representation of numbers replaced roman numerals not because Arabs conquered Romans, but because it works better. And it works better because it is more closely aligned with deeper mathematical structure: the same structures used to manipulate arabic-style digits are used to manipulate polynomials. This is why students taught to work with numbers using algorithms that cleanly reflect this structure find the transition to polynomials relatively painless. If number work obscures the structure (e.g. with certain addition tricks, or mechanical aids such as an abacus, slide rule or calculator) then students tend to see polynomials as a new and difficult subject. They have learned to deal with trees, but without absorbing the viewpoint needed to see the forest.

[2]The lack of agreement on meaning of terms is further explored in Communication between the mathematical and math–education communities

K-12 students are not tested on abstract structure so it is up to teachers to make sure structure is clearly reflected in the materials. Standards documents could help by making explicit the key abstract structures involved in a particular topic, pointing out where else these ideas appear in the curriculum, and offering sample problems that display the structure.

2.6. **Summary.** Ideally a standards document will provide an effective common reference point for all concerned parties. Specifically:

- Test constructors should see what sort of problems are appropriate, and further see how problems might probe absorption of underlying mathematical structure.
- Teachers should be able to anticipate tests in detail, and see the underlying structures (general principles) the tests are supposed to support. Ideally the document should be more useful for this than a test derived from it.
- Teachers should see with some precision what the students have already done in earlier courses and so should not be repeated in non-remedial courses.
- Teachers should see how skills to be acquired in their course—and ways of thinking underlying these skills—will be needed later. More generally the document should coordinate connections between the material and the structure of mathematics.
- Textbook writers should see how to expand the material in ways useful to students and teachers.
- The focus should be on testable outcomes and content. Methods used to achieve these outcomes should be left to teachers, curriculum developers, and other education professionals. In particular the document should not be, or resemble, a set of lesson plans.

Unfortunately most standards documents are developed in politicized and often contentious processes that overlook most of these points and cannot address any of them effectively.

3. More About Tests

High-stakes tests provide enforcement and accountability for the implementation of standards. They are intended to powerfully influence learning so great care ought be taken to ensure this influence is beneficial. As with so much else this is a job that by default falls to the standards document.

3.1. **Tests as instruments of terror.** System-wide tests are typically given once, though a few systems have “second chance” administrations. Stakes are high for both students and teachers so teachers (and occasionally parents) emphasize this to motivate students to prepare. Stress levels are high. Test formats, grading criteria, and even question types are different from those typically used in class and this is another source of confusion and stress.

In these circumstances strong students usually do consistently well and weak students do consistently poorly. Outcomes for average students tend to be less reproducible: repeated tries at equivalent tests give scores with significant spread that seems random. Any given score doesn’t correlate well with anything, so in particular cannot correlate with learning.

In some communities there is vocal opposition to high–stakes testing, and the drawbacks noted above are often given as reasons to end it. The reality is that testing is here to stay, but this does not mean the drawbacks are not real or not important. Problems should be honestly acknowledged and fixing them should be an urgent priority for test designers and administrators. However it is hard to imagine how this could happen without guidance from the standards document.

3.2. **Tests as defective standards.** Tests are traditionally thought of as assessment instruments and not part of the educational process itself.

When there is a lot of material assessment tests usually spot–check at random: if the student does not know what will be omitted then comprehensive learning is needed for reliable good performance. Similarly if generic problems are time-consuming then tests may use artificially simplified cases. If the student does not know how problems may be simplified then again comprehensive learning is needed for reliable good performance.

The traditional disconnect between tests and learning does not hold for system-wide high-stakes tests. Old tests are available and carefully scrutinized and new ones are expected to be consistent with them. Tests become de facto standards so simplifications or omissions are incorporated in and weaken the curriculum. The converse to this is that to avoid weakening the curriculum standard tests would have to be harder and more comprehensive than they are now. Clearly our approach to testing must change dramatically if this is not to be a prescription for massive failure.

A related problem is that current multiple-choice tests tend to drive curricula away from abstract and symbolic work. Symbolic expressions have structure that may give shortcuts to identification of correct answers and it is common practice to hide such structure by numerical evaluation. πr^2 for instance is instantly recognizable as the area of a circle, while 16.6 is not obviously the (approximate) area of a circle of radius 2.3. This leads to high-stakes tests dominated by approximate numerical problems, and this in turn de facto establishes the goal of the course as success with numerical problems. Students come out knowing exactly what to do with a problem involving a circle of radius 2.3, but are stumped by the same problem when the radius is given as "r". This is particularly acute in curricula emphasizing use of calculators. These students have missed the benefits of math as an introduction to abstract logical reasoning, and are at a disadvantage in college courses.

3.3. **Tests as suppressors of quality and diversity.** In the last 40 years the US K–12 system improved in some ways, going from one in which many children dropped out to one with a realistic hope that none need be left completely behind. However there have been costs including a decline in achievement levels: since the priority is now to get everyone over one bar it has to be set low. Resources are focused on weak students since they are at risk of failure and good students are not; a great shift from the Sputnik era goal of boosting the best.

Declining preparation of high–school graduates has driven a corresponding decline in post–secondary achievement and American students have nearly disappeared from top achievement levels. Our better graduate schools are populated by high–achieving students from other countries and our leading scientists, engineers, and educators are increasingly international. Significant parts of our way of life are now maintained by importing high-quality K–12 and undergraduate education.

Dependence on foreign educational systems for high–quality preparation is a threat to our national security and prosperity. Eventually we must do better with our own good students. Any real movement in this direction would have to be supported—if not started—by standards documents and system-wide tests.

For example a state might have two levels of tests, say "general" and "college prep". A bad score on the college prep test could be converted to a good score on the general test, so no one would "fail" college prep. College prep tests and standards would organize development of more-demanding courses and therefore increase diversity in the system.

It should be emphasized that the need is for better preparation in high–school subjects such as algebra, geometry and trigonometry, *not* in topics such as calculus and statistics. Very few schools have the resources to do a college–quality job with college–level subjects. Mediocre or mechanical courses (driven for instance by the AP calculus test) give little advantage to college students, and certainly do not make up for weak preparation in algebra and geometry.

3.4. **Summary.** Consistency is the overwhelming concern in traditional high–stakes test design. Tests must be similar in content and scores should be as consistent as possible from one administration to the next. This is difficult and expensive but test developers do impressively well at it. In contrast current tests show little or no evidence of concern for the effect they have on the instructional program. It may be that design criteria and pressures during the development process make this impossible, but the end result is a consistently negative influence and no reason for hope that traditional approaches to test construction will produce anything else.

4. Conclusion

The 2004 version of this article went on to suggest ways to make standards documents more functional. However at the time of the revision in late 2008 the NCTM and many states have revised their documents and other organizations including the College Board have issued Standards, all perpetuating the defects discussed above. The federal No Child Left Behind regulations has further polarized and obscured many issues. The National Mathematics Advisory Panel identified a few of the problems but was unable to come to a firm conclusion on most of them. It no longer seems reasonable to hope for significant change in the way Standards Documents are constructed.

It seems remotely possible that good tests could be developed outside the Standards system, see Beneficial High–Stakes Math Tests: An Example.