Title

Summative and Formative Assessments in Mathematics Supporting the Goals of the Common Core Standards

Journal

Theory Into Practice, 54(3)

ISSN

0040-5841

Author

Schoenfeld, Alan H

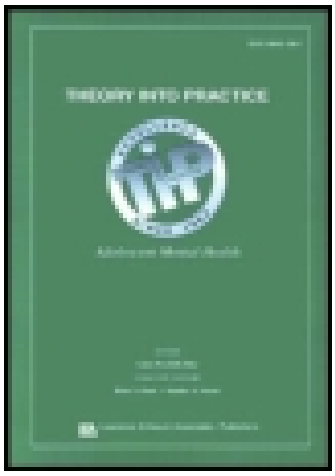
Publication Date

2015-07-03

DOI

10.1080/00405841.2015.1044346

Peer reviewed



Theory Into Practice

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/htip20

Summative and Formative Assessments in Mathematics Supporting the Goals of the Common Core Standards

Alan H. Schoenfeld

Accepted author version posted online: 05 May 2015.

To cite this article: Alan H. Schoenfeld (2015): Summative and Formative Assessments in Mathematics Supporting the Goals of the Common Core Standards, Theory Into Practice, DOI: 10.1080/00405841.2015.1044346

To link to this article: http://dx.doi.org/10.1080/00405841.2015.1044346

Alan H. Schoenfeld

Summative and Formative Assessments in Mathematics

Supporting the Goals of the Common Core Standards

Alan H. Schoenfeld is the Elizabeth and Edward Conner Professor of Education and affiliated professor of Mathematics at the University of California, Berkeley.

Correspondence should be addressed to Professor Alan Schoenfeld, Education, EMST, Tolman Hall #1670, University of California, Berkeley, California 94720-1670. Email: alans@berkeley.edu.

Being proficient in mathematics involves having rich and connected mathematical knowledge, being a strategic and reflective thinker and problem solver, and having productive mathematical beliefs and dispositions. This broad set of mathematics goals is central to the Common Core State Standards for Mathematics.

High stakes testing often drives instructional practice. In this article I discuss test specifications and sample assessment items from the two major national testing consortia and the prospects that their assessments will be positive levers for change.

For more than 20 years the Mathematics Assessment Project has focused on the development of assessments that emphasize productive mathematical practices, most recently creating "Formative Assessment Lessons" (FALs) designed to help teachers build up student understandings through focusing on student thinking while engaging in rich mathematical tasks. This article describes our recent work.

Introduction

The United States stands at the threshold of significant changes in mathematics assessment, both in terms of what kinds of understandings are assessed and in terms of the increasing homogeneity of mathematics assessments, nationwide. These changes reflect the continued evolution of the "standards movement", which can be dated back to the of the National Council of Teachers of Mathematics' (NCTM, 1989) production of the *Curriculum and Evaluation Standards for School Mathematics* in 1989 and to a radical change in the national high stakes accountability context due to the "No Child Left Behind" legislation passed by Congress in 2001 (U.S. Government Printing Office, 2002). Within a few years the vast majority of American students will be taking one of 2 high stakes examinations, both of which are intended to represent the mathematical values represented in the Common Core State Standards for Mathematics, or CCSSM (Common Core State Standards Initiative, 2010). To the degree that the assessments represent the values in CCSSM, and to the degree that high stakes assessment drives instruction, mathematics teaching in the US will be much more focused and coherent than it has been over the past quarter century.

In what follows I focus on 2 kinds of assessments: *Summative assessments* are examinations or performance opportunities whose primary purpose is to assign students a score on the basis of their knowledge, such as end-of-course exams, SATs, or state or national high stakes exams. *Formative assessments* are examinations or performance opportunities whose primary purpose is to provide student and teachers feedback about the student's current state while there are still opportunities for student improvement (see, e.g., Black & William, 1998a, 1998b, 2009; *Educational Designer* special issue, October 2014; Hernandez-Martinez, Williams, Black, Davis, Pampaka, & Wake, 2011).

This introduction briefly describes the evolution of mathematics standards and the national testing context. I then examine some typical current test items, and some of the items that represent the assessments being produced by the 2 national assessment consortia, the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC). Issues of alignment with the CCSSM remain; but, assuming that these can be worked out, the new assessments portend significant change. The question, then, is how to prepare students and teachers for such change. I describe one attempt, a series of Formative Assessment Lessons (FALs) created by the Mathematics Assessment Project ((Mathematics Assessment Project, 2014).

The Evolution of Standards, 1975-2010

Prior to 1989, mathematics curriculum documents focused almost exclusively on the mathematical *content* (e.g., operations on numbers; measurement; algebra; geometry) that students were to learn. This changed when the national Council of Teachers of Mathematics' (1989) *Curriculum and Evaluation Standards for School Mathematics*, reflecting current

research, emphasized cross-cutting *processes* of doing mathematics: problem solving, reasoning, communicating with mathematics, and making connections using mathematics. This trend continued, with NCTM producing an updated version of standards in 2000 (NCTM, 2000), and with groups like the National Research Council (2001) painting the picture of mathematical proficiency reflected in Figure 1. The core idea is that conceptual understanding and procedural fluency, the main foci or prior instruction, are not enough; true mathematical proficiency also includes developing a positive disposition toward mathematics, the ability to approach new problems and use the knowledge one has developed in other contexts, and to do so strategically.

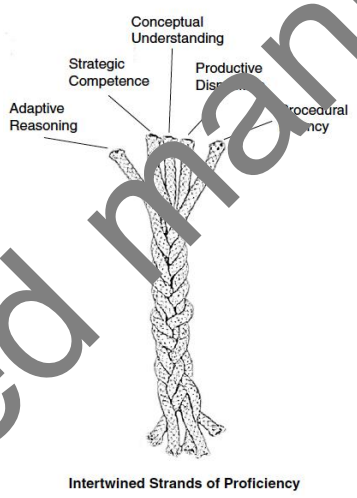


Figure 1. The representation of mathematical proficiency in *Adding it Up (National Research Council, 2001, p. 5)*

The *Common Core State Standards* represent the natural evolution of these ideas. They provide content specs at each grade level, with an emphasis on the focus and coherence of the mathematics to be learned. And, an emphasis continues on how students are to engage with mathematics, now referred to as "Standards for Mathematical Practice." The 8 mathematical

practices highlighted in the CCSSM (see specifically pages 6-8 of CCSSM, 2010) are that students will

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to Precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

The challenge for assessment has been, and will continue to be: Is it possible to assess student performance of such practices in ways that are reliable and valid? (see, e.g., Delandshere & Petrosky, 1998.)

The Curriculum and Assessment Context, 1975 - present

In the 1970s and through the 1980s, a small number of states had statewide mathematics standards; a smaller number (e.g., California, New York, and Texas) had assessments that were aligned to those standards. In effect, each state was free to do what it wanted with regard to curriculum and assessment – within the bounds of college requirements, standardized tests, etc. Substantial variation existed across states until the passage of the "No Child Left Behind" Act of 2001. To qualify for federal funding under NCLB, as it is known, each of the states had to institutionalize standards for mathematical performance, and to assess students on a regular basis. These exams were *high stakes*: students' promotion, teachers' salaries (and jobs),

administrators' salaries (and jobs), and the very existence of schools and districts (which could be dismantled if student test scores failed to meet the increasingly stringent scoring requirements over a period of years) depended on test scores. The result was to distort the system, where many teachers and districts did whatever was necessary to score well. Not surprisingly, most schools focused heavily on teaching to the tests, which were of highly variable quality. Given that each state had its own standards and assessments, the result was nationally institutionalized incoherence (See, e.g., Azzam, Perkins-Gough, & Thiers, 2006).

This began to change with the US Department of Education's (2009) Race to the Top (RTT) program, announced by President Obama and Secretary of Education Duncan on July 24, 2009. The constraints of RTT were that *consortia* of states, not individual states, would apply for funding. This constraint led the Council of Chief State School Officers and the National Governors Association to sponsor the Common Core State Standards Initiative, which produced the CCSSM. As of this writing 43 states, the District of Columbia, 4 territories, and the Department of Defense Education Activity (DoDEA) have adopted the Common Core State Standards – thus establishing what is a *de facto* set of national mathematics standards.

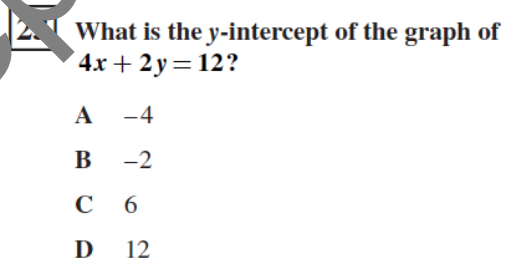
In addition, the Race to the Top Assessment Program "provided funding to [two] consortia of States to develop assessments that are valid, support and inform instruction, provide accurate information about what students know and can do, and measure student achievement against standards designed to ensure that all students gain the knowledge and skills needed to succeed in college and the workplace." (US Department of Education, 2013, p. 1) Those consortia, the Partnership for Assessment of Readiness for College and Careers, or PARCC, and the Smarter Balanced Assessment Consortium, or SBAC, have between them enrolled the majority of the states that have agreed to align themselves with the Common Core State

Standards. Other states are producing their own assessments, which are intended to be aligned with the Common Core – as opposed to being aligned with their previous state standards. As a result, a patchwork of 50 state assessments will no longer exist. The vast majority of students across the country will be faced with one of 2 assessments, constructed either by PARCC or SBAC, and ostensibly aligned with the CCSSM. Given WYTIWYG, and the fact that CCSSM standards and assessments will be given at each grade K-8, there will be a degree of homogeneity in curricula and in assessments that is unprecedented in American history.

The Nature of Mathematics Assessments, Past and Possible Future

Mathematics assessments across the US have varied widely from state to state. Here I provide an example from the California Standards Tests (CSTs) as an example of what has been the reality in one state, and contrast this with a richer assessment of proficiency in the same content area. I then discuss the item specifications and sample items from the 2 national assessment consortia.

Figure 2 contains a representative eighth grade algebra problem from the CST.



What is the *y*-intercept of the graph of $4x + 2y = 12$?

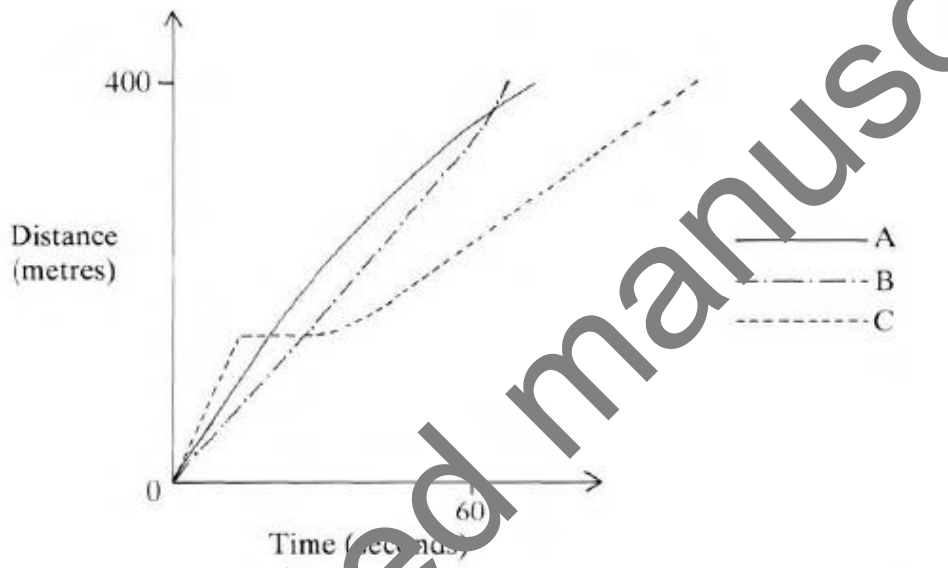
A −4
B −2
C 6
D 12

Figure 2. A released CST problem from the 8th grade algebra I test

This task, like most of those on the CST, focuses on content knowledge. There are at least 3 straightforward ways to get the answer: by substituting $x = 0$ into the equation and solving

the resulting equation, $2y = 12$; by writing the equation in the slope-intercept form $y = -2x + 6$; and by writing it in the 2-intercept form $x/3 + y/6 = 1$. In each case, the procedure is mechanical and the answer straightforward to obtain. Although content knowledge is assessed, it is hard to argue that the standards for mathematical practice are assessed in any meaningful way.

In contrast, consider the "hurdles race" task given in Figure 3.



The rough sketch graph shown above describes what happens when 3 athletes A, B and C enter a 400 metre hurdles race.

Imagine that you are the race commentator. Describe what is happening as carefully as you can. You do not need to measure anything accurately.

Figure 3. Hurdles Race. Swan, M., and the Shell Centre Team (1985), p. 42. Reprinted with permission.

This question calls for interpreting distance-time graphs in a real-world context, a central component of mathematical modeling. A complete response includes

- Understanding that a runner whose graph appears "to the left" of another is *ahead* at that point, having taken less time to travel the same distance. (Thus B wins the race);
- Understanding what points of intersection signify in this context (that 2 runners have run the same distance at the same time, so they are tied at that point in the race);
- Interpreting the horizontal line segment (the runner is not progressing, so – in the context of a hurdles race – must have tripped on a hurdle and fallen), *and*
- Putting all of the above together in a coherent narrative.

Equally important, responding appropriately to this question calls for demonstrating proficiency at (at least) the first 4 of the mathematical practices highlighted above: The students have to persevere in sense making and problem solving and reason abstractly and quantitatively, constructing reasoned explanations of "real world" phenomena. *If* tasks of this level of complexity will appear on the 2 consortia's assessments, then there will be significant changes in what is assessed (and, by virtue of WYTIWYG, what is taught) across the nation.

Thus, there is significant promise that the 2 assessment consortia can move things in very productive directions – but progress is hardly guaranteed. There are various places where things can go wrong: in the specifications for the exams; in ways the specifications are realized in the exams themselves; and in the grading, to mention only 3.

The Consortia's Exam Specifications

Here I think there are grounds for significant optimism. The fundamental change in the SBAC assessments is that they will report either 3 or 4 scores, not just one. Until now, a student's score in most assessments was a number on a given scale – so many points out of 100 on some tests or, say, a numerical score between 200 and 800 on the SAT. (See

http://sat.collegeboard.org/scores/understanding-sat-scores for a description of how to interpret such scores.). Such reporting provides an indication of how well the student did, but it provides no information about what the student did or did not do well. (For example, did the student do well on algebra but not geometry, or vice-versa? Did he or she earn most of his points on procedural questions, on those that asked for extended chains of reasoning, or on some of both?) In contrast, the SBAC (2012, p. 19) test specs call for reporting 4 scores for each student, corresponding to knowledge of: *Concepts & Procedures; Problem Solving; Communicating Reasoning;* and *Modeling and Data Analysis.* If the tests and the reporting provide meaningful opportunities to demonstrate proficiency in these areas, this will broaden instructional foci in desirable ways. (And, this will open up room for meaningful formative assessment, as described below.)

This is promising. It is quite clear that a test like the California Standards Test, with only multiple-choice problems focusing on concepts and procedures, fails to assess claims 2, 3, and 4 in a meaningful way. Extended problem-solving tasks, of complexity not unlike the "hurdles race" task, populate the SBAC specifications. If such tasks make their way into the actual assessments, they will (by virtue of WYTIWIG) drive classroom instruction in the direction of the CCSSM. But there are risks.

The PARCC assessment promises tasks of 3 types: (1) Tasks assessing concepts, skills and procedures, (2) tasks assessing expressing mathematical reasoning, and (3) tasks assessing modeling/applications (PARCC, 2012, p. 14). This is broadly consistent with the approach taken by SBAC and the CCSSM. It is not clear from the documents available on the PARCC website (http://www.parcconline.org/) what the format for reporting student scores will be, so I was unable to determine whether there will be separate scores for the 3 categories listed above. If

there is only a single score, it will be difficult for users (including teachers) to know where to focus their attention when preparing for the tests.

The Consortia's Plans for Scoring

A major challenge that the consortia face is how to score of millions of students' tests in a relatively short time frame (a matter of weeks). Here we are in somewhat unknown territory, and I find the prospects troubling. SBAC plans to use a significant amount of computer-adapted testing; the PARCC assessments will be administered via computer, and a combination of automated scoring and human scoring will be employed" (http://www.parcconline.org/parcc-assessment-design).

I have several concerns with computer-based "efficiency." The goal of both consortia is to move toward all assessments being given only on computers, and being completely computer-scored. I am far from convinced that the state of the art with regard to the computer grading of "essay questions" in mathematics – especially those that employ diagrams and other mathematical representations – can deliver the accurate assessment of student work that is needed. As it stands now, creating diagrams on available interfaces is a clumsy and time-consuming process (something I can sketch in 30 seconds can take more than a few minutes to produce on a computer screen), and I have yet to see programs that could do a good job of scoring student responses to problems like the one given in figure 3.

I have equally large reservations about the very notion of computer-adaptive scoring. Such scoring may be appropriate when the goal is to simply assign one score, and reporting on content and practices is not central. (That should not be the case here!) But worse, students who get off to a shaky start by giving the wrong answers to the first 2 problems on a test with

computer-adaptive scoring may never have the opportunity to demonstrate what they know. The primary determinant of the "next" questions in computer-adaptive testing is item difficulty, the goal being to converge rapidly on a student score. This may be efficient, but it does not serve the needs of students or teachers by providing information about what students know and can d

Sample released Items

It is mid-2014 as I write this article, and the situation is in flux. Both consortia are pilot-testing their exams as I write, with the expectation that the assessments will be used for the 2014-2015 academic year. Sample items are now available on the PARCC and SBAC web sites; see, for example, <http://www.parcconline.org/samples/item-task-prototypes> and < http://sampleitems.smarterbalanced.org/itempreview/sbac/index.htm>. Readers should review the items and form their own opinions. Overall, the released items suggest some, but not much, reason for optimism. Some items make good use of the technology, even at the "basic knowledge" level; consider for example, SBAC item 42960 (Figure 4). The computer-based format improves on the "matching" format used in many paper and pencil or computer tasks.

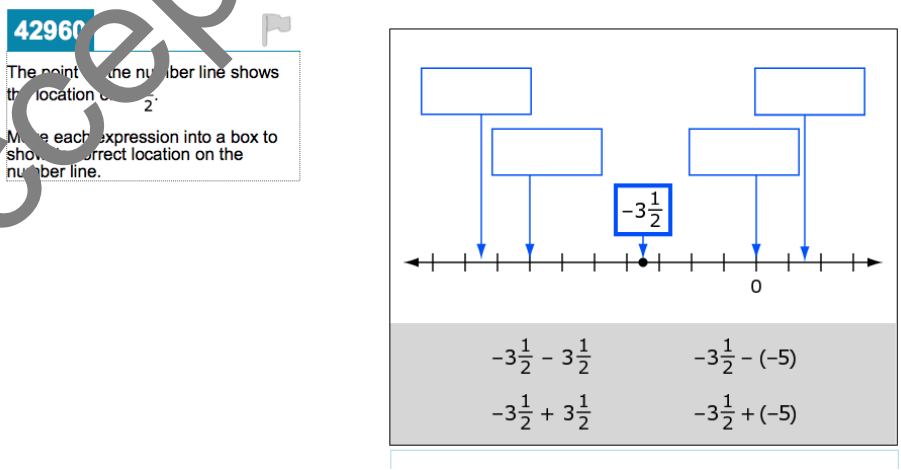


Figure 4. A computerized version of a "matching" problem.

I am less sanguine about some of the open-ended questions on both assessments. Given time constraints (not much time is allocated for open questions, so there may be only one or 2 per assessment) and the challenges of scoring such exams via computer, the current exemplars may move mathematics assessment significantly forward. One can hope that the exams will evolve over time.

Formative assessment

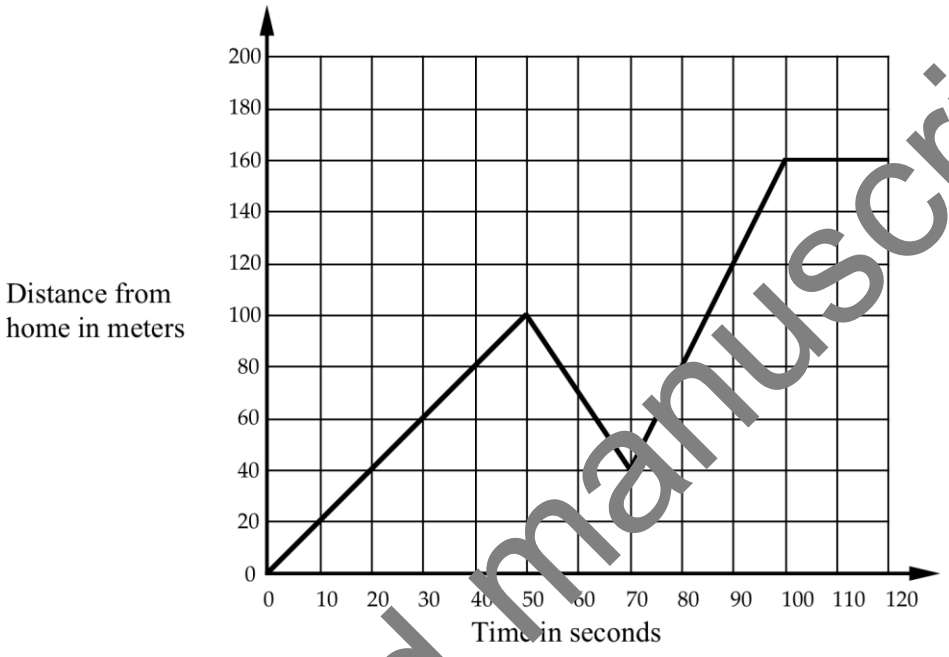
A major challenge facing teachers, especially those whose instructional focus has primarily been on procedural items such as the one in Figure 2, is to help students develop the skills and understandings required to address tasks like the one in Figure 3. Part of that challenge is dealing productively with student approaches – both correct and incorrect – as students grapple with complex tasks. One prevalent approach is using *formative assessment,* which provides information about student understanding at a point when the teacher and students can act productively on that understanding, rather than demonstrating what students "know and can do" *after* instruction (See Black & William, 1998, for a classic overview).

The Mathematics Assessment Project (MAP), for which I am Principle Investigator, has been producing formative assessment lessons (FALs) intended to support teachers in conducting formative assessments. As I write, nearly 100 FALs are available for free on the MAP web site, <http://map.mathshell.org/materials/index.php>. To convey the flavor of the approach taken by the project, I briefly describe the FAL "Interpreting distance-time graphs," <http://map.mathshell.org/materials/lessons.php?taskid=208&subpage=concept>.

FALs begin with a diagnostic problem that the students work *before* the lesson, so that the teacher is provided information about the students' likely strengths and pitfalls. The diagnostic problem for "interpreting distance-time graphs" is given in Figure 5.

Journey to the Bus Stop

Every morning Tom walks along a straight road from his home to a bus stop, a distance of 160 meters. The graph shows his journey on one particular day.



1. Describe what may have happened.
   You should include details like how fast he walked.
2. Are all sections of the graph realistic? Fully explain your answer.

Figure 5. Diagnostic problem for "interpreting distance-time graphs."

The FAL lesson plan suggests that the teacher respond to the student work not by assigning scores, but instead by creating a set of questions that address the issues revealed by what the students have written. It supports the teacher as it identifies typical student misinterpretations and suggests questions that might push student thinking further. Common issues include (a) Student interprets the graph as a picture; (b) Student interprets graph as speed–

time; (c) Student fails to mention distance or time; (d) Student fails to calculate and represent speed; (e) Student misinterprets the scale; and (f) Student adds little explanation as to why the graphs is realistic. A sample set of questions for issue (a) is given in Figure 6.

Common Issue	Suggested Questions and Prompts
Student interprets the graph as a picture For example: The student assumes that as the graph goes up and down, Tom's path is going up and down. Or: The student assumes that a straight line on a graph means that the motion is along a straight path. Or: The student thinks the negative slope means Tom has taken a detour.	•If a person walked in a circle around their home, what would the graph look like? •If a person walked at a steady speed up and down a hill, directly away from home, what would the graph look like? • In each section of his journey, is Tom's speed steady or is it changing? How do you know? • How can you figure out Tom's speed in each section of the journey?

Figure 6. A sample student issue and questions to explore it.

The goal is for the teacher to annotate the student work (individually if time permits, or by way of a list of "thought questions" for the class if not), so the students can engage more fully with the content. The full 90-minute lesson begins with a whole-class discussion of the problem in Figure 7. The students are asked to decide which of the stories A, B, and C corresponds to the distance-time graph that appears in the figure, and a whole-class discussion of the reasons students had for their choices follows. The result of this discussion is an annotated graph, which looks something like Figure 8.

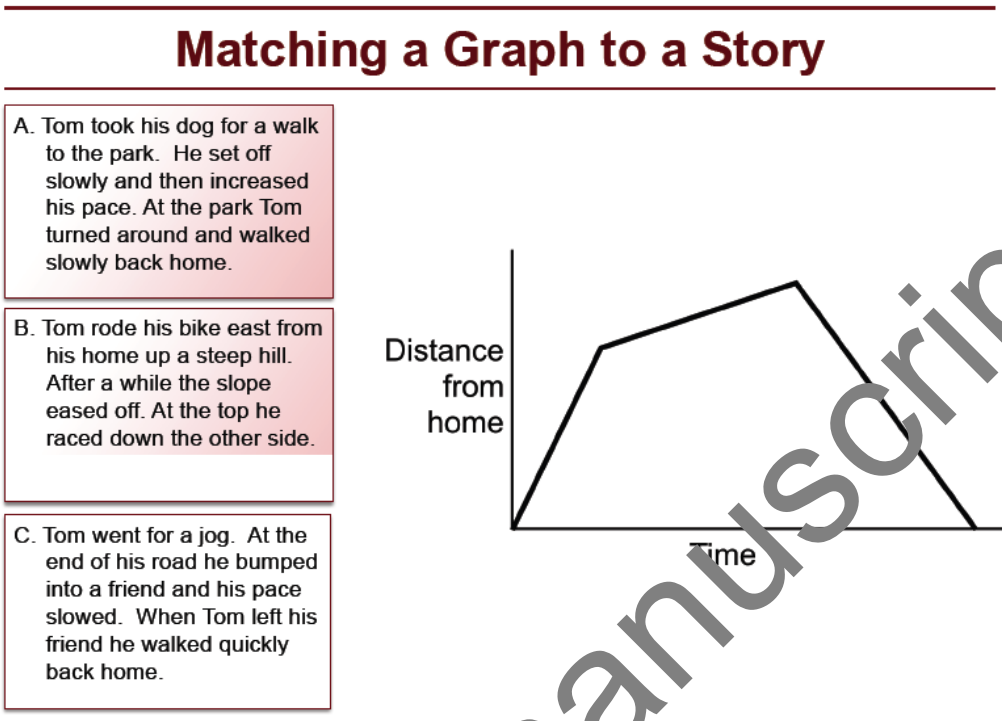


Figure 7. A distance-time graph question to start the lesson.

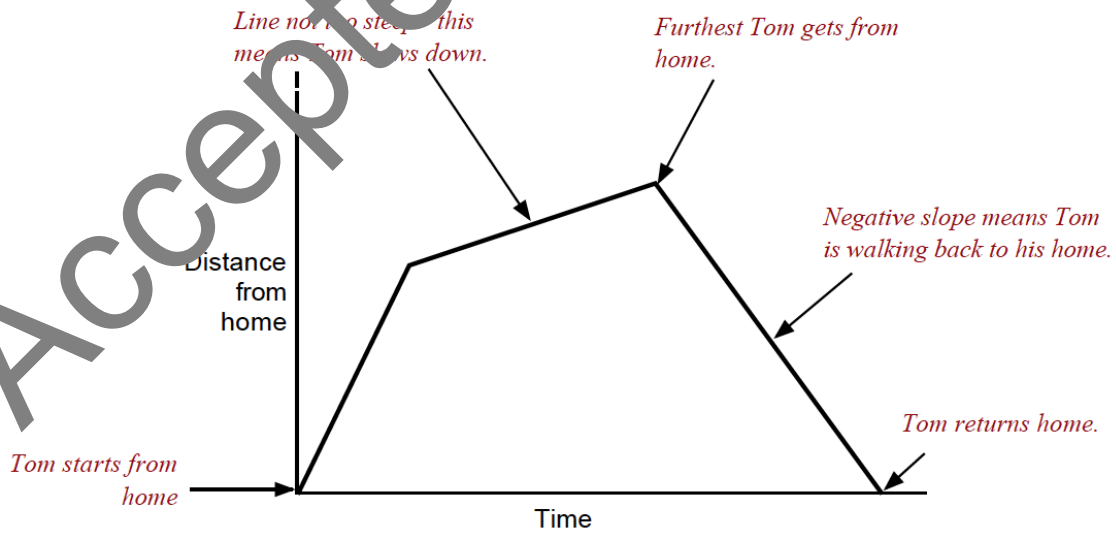


Figure 8. An annotated graph.

With this as backdrop, the main part of the lesson, a card-matching exercise begins. Students are given a set of 10 distance-time graphs and 10 stories. They are asked to work in small groups, matching the stories to the graphs. A sampling of the first 4 distance-time graphs is given in Figure 9.

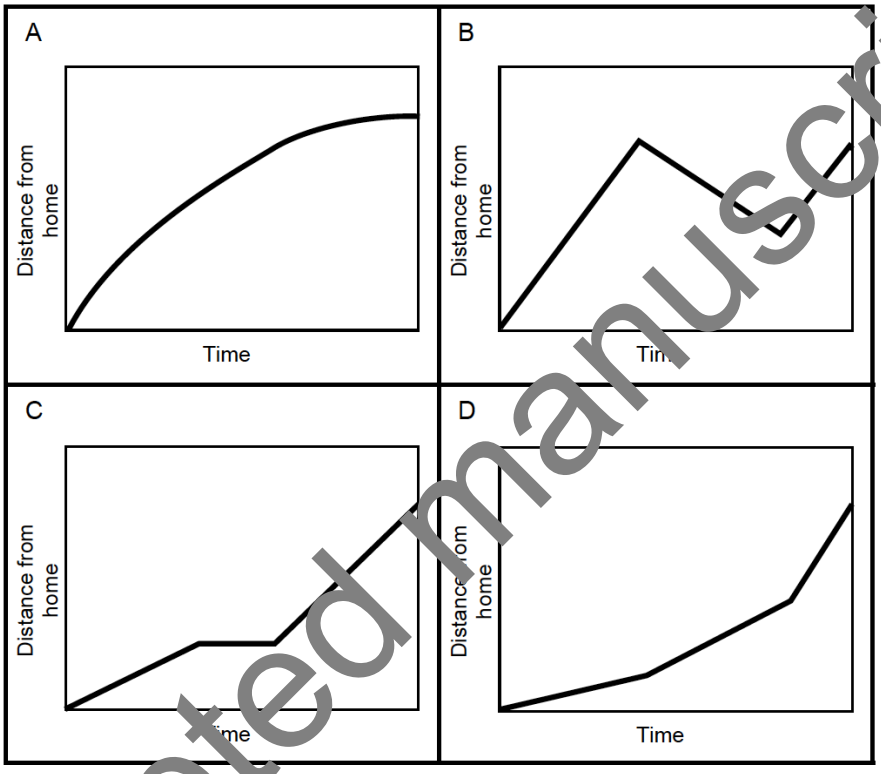


Figure 9. Sample distance-time graphs for the card sort.

Four of the 9 filled-out stories are shown in Figure 10. The tenth card says, "Make up your own story."

1 Tom ran from his home to the bus stop and waited. He realized that he had missed the bus so he walked home.	**2** Opposite Tom's home is a hill. Tom climbed slowly up the hill, walked across the top, and then ran quickly down the other side.
3 Tom skateboarded from his house, gradually building up speed. He slowed down to avoid some rough ground, but then speeded up again.	**4** Tom walked slowly along the road, stopped to look at his watch, realized he was late, and then started running.

Figure 10. Sample stories for the card sort.

As students work on the sorting task, they often encounter untenable situations – e.g., they have 2 incommensurate stories for the same graph or 2 different graphs for the same story. This gives rise to heated conversations about *why* stories and graphs do or do not match.

At this point in the lesson, the teacher, who has been monitoring the discussions, starts a conversation about how to resolve the conflicts. He or she introduces the idea of building a table from a graph, as illustrated in Figure 11.

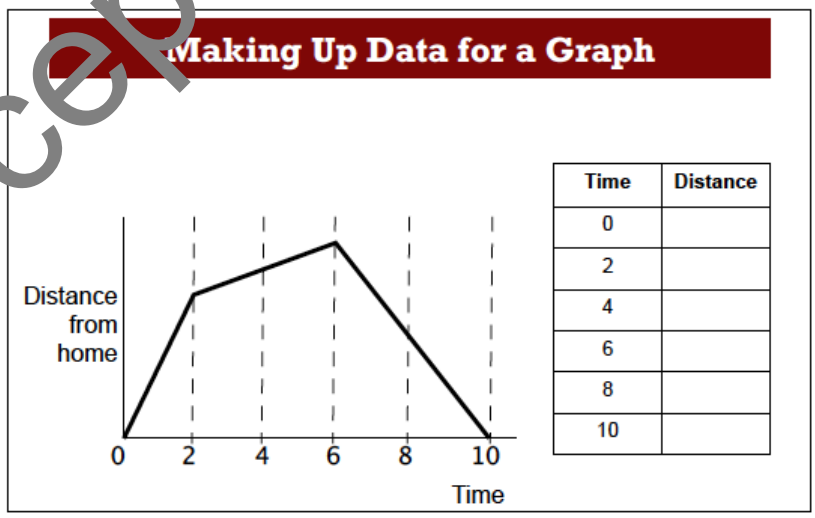


Figure 11. Building a table from the graph.

The students are then given a third set of cards, which contains a collection of distance-time tables. Their task now is to use the tables to reconsider their graph-to-story pairings, and to put together a poster that features ten matching triples, each containing a story, graph, and table that are mutually consistent. The students share their posters, compare and contrast results in a group. The lesson ends with students being given time to revise their posters on the basis of what had been discussed during the whole class discussion.

I note, briefly, that this kind of lesson supports student engagement in a number of fundamentally important aspects of learning: dealing with conceptually rich mathematics, being given the opportunity to engage (and be supported in engaging with) challenging problems, and to discuss and present their own ideas. My research group has been developing a set of tools for supporting classroom activities of this type. See Schoenfeld (2014) and the "TRU Math" suite of tools at <http://map.mathshell.org/materials/trumath.php>.

Discussion

The United States stands at a crossroads with regard to mathematics education, with assessment playing a major role and a potential lever for change. The potential for significant change comes with (a) the adoption of the CCSSM by the vast majority of states, and (b) the fact that most of the states that have aligned with the CCSSM will be using one of only 2 assessments (PARCC or SBAC) to assess student proficiency. Condition "a" suggests that we will have, for the first time in the US, a de facto national curriculum. Condition "b" suggests that the 2 current assessments, because of the high stakes involved, will play a fundamental role in shaping how that curriculum comes to life in American classrooms. If the assessments focus on the mathematical values intended in the CCSSM, a great potential for assessment-driven progress exists; but if the assessments pervert the mathematical intentions of the CCSSM writers for

reasons of cost, ease in scoring, or psychometric considerations, the results can be disastrous. The stakes are indeed high. This is the time for a serious investment in an R&D agenda, so that the system can be self-improving.

The right assessments can orient the system in the right directions, but even so, there are issues of system capacity. Teaching for the kinds of content understandings and mathematical practices described in the CCSSM is hard. Generally speaking, teacher preparation programs have not had the time or resources to help teachers become proficient at formative assessment; nor does the current generation of texts provide teachers with adequate support. Formative assessment, well done, can support teachers in building rich mathematical classroom environments. It is our hope that the kinds of FALs described in this paper will help to provide such support.

Notes

1. No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425-2094, 2002.
2. Some years ago Hugh Burkhardt coined the phrase "What You Test Is What You Get (WYTIWYG)" to represent this reality. Space does not permit a discussion of WYTIWYG; see Barnes, Clarke, & Stephens (2000) and Bell & Burkhardt (2001).
3. See http://www2.ed.gov/programs/racetothetop/index.html.
4. See http://www.corestandards.org/standards-in-your-state/. These figures downloaded December 7, 2014.
5. See http://www2.ed.gov/programs/racetothetop-assessment/index.html.
6. See http://www.parcconline.org/about-parcc.
7. See http://www.smarterbalanced.org/.

8. This is problem 23 from the Algebra I released problems from the California Standards test, downloadable from the California Department of Education at <http://www.cde.ca.gov/search/searchresults.asp?cx=001779225245372747843:gpfwm5rhxiw&output=xml_no_dtd&filter=1&num=20&start=0&q=released%20items%20california%20standards%20test>. If is typical of the level of difficult of the exam. More sample questions can be accessed at <http://starsamplequestions.org/starRTQ/search.jsp>.
9. Full disclosure: I was lead author for the SBAC mathematics content specifications.
10. http://www.smarterbalanced.org/smarter-balanced-assessments/.
11. It is absolutely essential for the mathematical integrity of the standards to drive the test construction process, with psychometric considerations then taken into account, rather than – as is typical in test construction – the other way around.

References

Azzam, A. M., Perkins-Gough, D., & Thiers, N. (2006). The impact of NCLB. *Educational Leadership, 64*, 94-96.

Barnes, M., Clarke, D., and Stephens, M. (2000). Assessment: the engine of systemic curricular reform? *Journal of Curriculum Studies*, *32*, 623-650.

Ben, A., & Burkhardt, H. (2001). Evidence on the influence of assessment on teachers' classroom practices. Downloaded October 31, 2001, from <http://www.toolkitforchange.org/toolkit/view.php?obj=1034&menu=i>.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in education, 5*, 7-74.

Black, P., & Wiliam, D. (1998b). *Inside the black box: raising standards through classroom assessment.* London, UK: King's College London School of Education.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment and Evaluation, 21*, 5-31.

California State Department of Education. Standardized Testing and Reporting - STAR Sample Test Questions. Downloadable at < http://starsamplequestions.org/starRTQ/search.jsp>.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics.* Downloaded July 9, 2010, from http://www.corestandards.org/the-standards.

Delandshere, G. & Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher, 27*, 14-25.

Educational Designer (2014, October). *Special Issue on Formative Assessment.* Downloaded November 1, 2014 from http://www.educationaldesigner.org/ed/volume2/issue7/.

Hernandez-Martinez, P., Williams, J., Black, L., Davis, P., Pampaka, M., & Wake, G. (2011). Mathematics coursework as facilitator of formative assessment, student-centred activity and understanding. *Research in Mathematics Education, 13*, 197-212.

Mathematics Assessment Project. (2014) The TRU Math Document Suite. Downloaded April 9, 2014, from <http://map.mathshell.org/materials/trumath.php>.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Research Council (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

No Child Left Behind Act of 2001. US Government Printing Office Pub. L. No. 107-110, 115 Stat. 1425-2094, 2002.

Partnership for Assessment of Readiness of College and Careers. (2012). Model content frameworks: Mathematics grades 3–11, version 2.0, August 31, 2012 *(revised)*. Downloaded September 23, 2012 from <http://www.parcconline.org/parcc-model-content-frameworks>.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? Manuscript submitted for publication.

Smarter Balanced Assessment Consortium. (2012). Content Specifications for the Summative Assessment of the *Common Core State Standards for Mathematics*. Downloaded March 20, 2012, from <http://www.smarterbalanced.org/smarter-balanced-assessments/>.

Swan, M., & the Shell Centre Team (1985). *The language of functions and graphs.* Nottingham, UK: The Shell Centre of Mathematics Education.

US Department of Education. (2009). Race to the top fund. Downloaded January 10, 2010, from <http://www2.ed.gov/programs/racetothetop/index.html>.

US Department of Education. (2013). Race to the top program technical review process. Downloaded December 3, 2013, from <http://www2.ed.gov/programs/racetothetop-assessment/technical-review-process.pdf>.