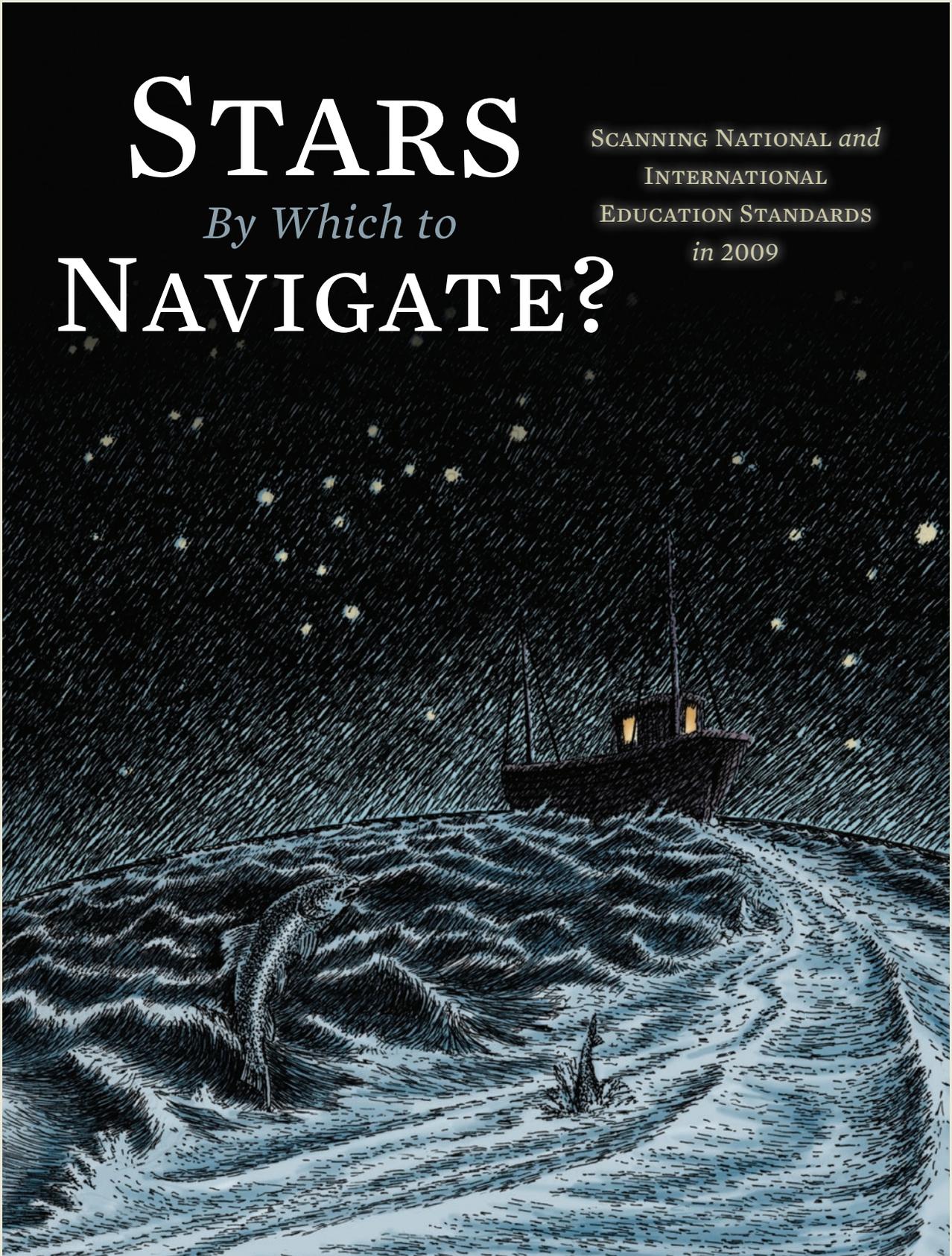


STARS

By Which to

NAVIGATE?

SCANNING NATIONAL *and*
INTERNATIONAL
EDUCATION STANDARDS
in 2009



AN INTERIM REPORT ON COMMON CORE, NAEP,
TIMSS AND PISA • OCTOBER 2009

Sheila Byrd Carmichael, W. Stephen Wilson, Chester E. Finn, Jr., Amber M. Winkler, and Stafford Palmieri

STARS

By Which to

NAVIGATE?

SCANNING NATIONAL *and*
INTERNATIONAL EDUCATION
STANDARDS *in* 2009

An Interim Report on Common Core, NAEP, TIMSS, and PISA

Sheila Byrd Carmichael, W. Stephen Wilson, Chester E. Finn, Jr., Amber M. Winkler, and Stafford Palmieri

Contents

1	Executive Summary • <i>Chester E. Finn, Jr., Amber M. Winkler, W. Stephen Wilson, and Sheila Byrd Carmichael</i>
4	Part 1: Introduction • <i>Chester E. Finn Jr. and Amber M. Winkler</i>
4	Introduction
8	Organization of the Report
8	Assumptions
10	Methods
11	Part 2: Grading
11	Math Content-Specific Criteria • <i>W. Stephen Wilson, with Richard Askey</i>
13	English Language Arts Content-Specific Criteria • <i>Sheila Byrd Carmichael, with Carol Jago</i>
16	Common Grading Metric • <i>Jointly developed by Fordham and the Reviewers</i>
19	Part 3: Reviews
19	Math • <i>W. Stephen Wilson</i>
20	—Common Core
22	—NAEP
26	—TIMSS
29	—PISA
33	English Language Arts • <i>Sheila Byrd Carmichael</i>
34	—Common Core
40	—NAEP
47	—PISA
51	Part 4: Closing • <i>Chester E. Finn, Jr. and Amber M. Winkler</i>
52	Part 5: Appendix • Overview of the Standards • <i>Stafford Palmieri</i>

Cover illustration by
Philipp Gasser

The Thomas B. Fordham Institute is a nonprofit organization that conducts research, issues publications, and directs action projects in elementary and secondary education reform at the national level and in Ohio, with special emphasis on our hometown of Dayton. It is affiliated with the Thomas B. Fordham Foundation, and this publication is a joint project of the Foundation and the Institute. For further information, please visit our website at www.edexcellence.net or write to the Institute at 1016 16th St. NW, 8th Floor, Washington, DC 20036. This report is available in full on the Institute's website. The Institute is neither connected with nor sponsored by Fordham University.

Executive Summary

■ SHEILA BYRD CARMICHAEL, W. STEPHEN WILSON, CHESTER E. FINN, JR., AND AMBER M. WINKLER

Subject-matter experts reviewed the content, rigor, and clarity of the first public drafts of the “Common Core” standards released in September 2009 by the Common Core State Standards Initiative (CCSSI) of the National Governors Association and Council of Chief State School Officers. Using the same criteria, the same experts also reviewed the reading/writing and mathematics frameworks of the National Assessment of Educational Progress (NAEP); the Trends in International Mathematics and Science Study (TIMSS); and the Programme for International Student Assessment (PISA). Letter grades were awarded to each.

The goal is to help U.S. educators and policymakers to judge the respective merits of these influential standards, de facto standards, and possible future standards. In particular, how do the draft Common Core standards stack up alongside extant national and international benchmarks?

Here are the grades:

- ◎ Common Core Reading/Writing/Speaking/Listening: **B**
- ◎ Common Core Mathematics: **B**
- ◎ NAEP Reading: **B** and NAEP Writing: **B**
- ◎ NAEP Mathematics: **C**
- ◎ TIMSS Mathematics: **A**
- ◎ PISA Mathematics: **D**
- ◎ PISA Reading: **D**

Here are the key findings, organized by framework:

COMMON CORE READING/WRITING/SPEAKING/LISTENING STANDARDS (SEPTEMBER 2009 DRAFT)

The drafters have done a praiseworthy job of defining essential competencies in reading, writing, and speaking and listening for success in both college and the workplace. They are also to be commended for not falling prey to spurious postmodern theories that disavow close reading and encourage interpretations of a text based solely on how it makes the reader *feel*. Further, the document properly acknowledges that essential communication skills must be embraced and addressed beyond the English classroom, which could lead to valuable collaboration among teachers and more consistent expectations across subjects.

These skill-centric standards do not, however, suffice to frame a complete English or language arts curriculum. Proper standards for English must also provide enough content guidance to help teachers instill not just useful skills, but also imagination, wonder, and a deep appreciation for our literary heritage. Despite their many virtues, these skills-based competencies cannot serve as a strong framework for the robust liberal arts curricula that will prepare young Americans to thrive as citizens in a free society. States adopting these standards must, therefore, be very careful about how they supplement them so as to achieve that goal.

COMMON CORE MATHEMATICS STANDARDS (SEPTEMBER 2009 DRAFT)

The draft covers most of the critical content and is coherent, organized, and clearly written. However, the standards are not explicit enough in how they address the arithmetic of rational numbers. Further, they do not set priorities, which means that readers are unable to discern which standards should be given more or less attention than others. As a result, many standards that will contribute little to college readiness are given equal status with standards that are essential foundations.

NAEP READING AND WRITING

NAEP's (separate but related) reading and writing frameworks function satisfactorily as the foundation for developing national, matrix-sample assessments of U.S. fourth, eighth, and twelfth graders' basic skills in these crucial areas. But as models for national and state standards they fall short of the mark.

The reading framework is generally clear and usable but could be more specific and detailed in ways that would also improve its content. The frameworks do not place enough emphasis on literature and English language conventions, nor do they distinguish the most desirable skill targets at each grade span. Making such revisions would help practitioners distinguish between more and less important content.

The writing framework also has shortcomings. Though the content is covered satisfactorily overall, more specificity at each level and more attention to writing conventions would help make the framework as rigorous as it should be. The guidelines for writing tasks are straightforward and coherent, and the sample prompts help to clarify expectations, but greater specificity at these important benchmark grades would also help indicate to educators what expectations should look like at intervening grades.

NAEP MATHEMATICS

Though the NAEP mathematics framework contains most of the important content we would expect students to learn, that content is hidden within an excessive number of standards (nearly 300). Less important material, such as data analysis, statistics, and probability, receives greater emphasis than more important material, like symbolic algebra. Throughout the document, important content standards are given the same status as secondary standards. Often standards containing crucial content are written in a way that also includes minor content, which makes it hard to identify the former. Over half of the standards would benefit from rewriting as they include unnecessary verbiage or mathematically meaningless language, or simply lack focus.

TIMSS MATHEMATICS

The TIMSS standards for both fourth and eighth grade math are clear, coherent, and well organized. They provide solid guidance to readers and users; they are measurable and burdened by very little jargon. They cover nearly all of the required content at appropriate levels of rigor. Excellent decisions have been made about what content to include and exclude, and there is little ambiguity about what is expected. Minor corrections to the pattern standards and inclusion of some missing standards would raise the score for TIMSS to a perfect ten.

PISA MATHEMATICS

The PISA mathematics framework and explanations do not cover the appropriate grade-level material and the released items indicate that the exam is quite weak in mathematical content. Further, PISA developers give only a vague description of what they mean by "mathematical literacy." This is a problem-solving test and, although mathematics is used, that seems almost incidental. Many problems have no apparent mathematical content. Because of this low level of required content knowledge, the claim that PISA tests "preparedness for further study" is rather dubious. Further, the test itself is unbalanced, overemphasizing data display. Most of the content that is expected of a 15-year-old on PISA is what younger students should have already mastered.

As a serious problem-solving test using elementary mathematics, the PISA assessment might function nicely. However, results from PISA ought not to be used to interpret how successful a school system is at imparting grade-level mathematical knowledge and understanding, nor are the PISA framework and released items a suitable model for U.S. standards setters at any level.

PISA READING

The PISA Reading Framework suffers from a number of problems. It does not address vocabulary development; offer specific expectations for reading and analyzing literary and non-literary texts; include specific expectations for the correct use of English language conventions; recognize the importance of literary heritage; or sufficiently define the amount, quality, and complexity of literary and informational texts that students should read. Much of the content focuses on metacognitive strategies for accessing information and addresses neither reading comprehension nor other critically important strands of English language arts that should be delineated in a set of strong national or state English standards for the U.S. (e.g., literature, writing, research, grammar and conventions, communication skills, etc.). Moreover, the framework is dense and murky, including multiple (and unnecessary) levels of detail about task types. Without any graphical explanation, it's difficult to follow and not helpful for informing curriculum and instruction.



These interim findings will be supplemented in spring 2010. Part II of this report will include science and history reviews as well as updated versions of some of the international frameworks and (provided these are available for review) appraisals of revised Common Core standards for end-of-high-school as well as earlier grades. Standards from the American College Testing Program (ACT), College Board, and the Progress in International Reading Literacy Study (PIRLS) will also be added.

Introduction

■ CHESTER E. FINN, JR. AND AMBER M. WINKLER

It might actually happen. Planets and stars are beginning to align. Dare we say it? Some sort of national education standards could become a reality.

If so, it will have been a long time in the making—at least half a century. In 1959, President Dwight D. Eisenhower called for “national goals” in education, including “standards.” A decade later, President Richard M. Nixon called “the fear of ‘national standards’” one of the “bugaboos of education.” In 1983, President Ronald W. Reagan accepted from his first education secretary *A Nation at Risk*, which sounded an alarm about the parlous condition of U.S. academic standards and arguably catalyzed twenty-five years of standards-based reform. In 1988, with the collaboration of the late Senator Edward M. Kennedy, President Reagan presided over the reinvention of the National Assessment of Educational Progress (NAEP), complete with state-by-state comparisons of student achievement and what became known as “achievement levels” by which NAEP data are now reported, a close relative of national standards.

In 1989, President George H.W. Bush assembled the governors in Charlottesville where they agreed to the first national education “goals” in U.S. history. He also paid for development of “voluntary national standards” in core subjects, only to see the Senate vehemently denounce the draft U.S. history standards—and most of the others crash and burn, too. President Bill Clinton later pushed for voluntary national testing, only to see a disgruntled House pull the plug on their funding. Then, in 2001, President George W. Bush and Congress teamed up to enact No Child Left Behind (NCLB), which embraced standards-based reform and testing—just not national standards and national tests. It turned out, however, that the hodge-podge of state standards and tests, some of them world-class rigorous, some downright embarrassing, made a mockery of NCLB’s 2014 drop-dead deadline by which all American youngsters would be “proficient” in reading and math.

September 2009 found President Obama’s Secretary of Education, Arne Duncan, also faulting NCLB for discouraging “high learning standards” and even “inadvertently encourag[ing] states to lower them.”

THE ADVENT OF COMMON STANDARDS

Though not everyone yet agrees, one solution now gathering force is development of a set of “common” core academic standards for the United States, much like most high-performing countries already have. That’s a heavy lift, to be sure, but the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO), in partnership with Achieve, ACT, and the College Board, have embarked on just such an undertaking.

Known as the “Common Core State Standards Initiative (CCSSI),” the goal of this state-led initiative is to develop common standards in reading/writing/listening/speaking and mathematics for grades K-12. (Science may eventually follow.) Governors and commissioners of education from nearly every state have already pledged some level of commitment to this initiative though no one’s yet quite sure what that means.

In mid-September, CCSSI released a “public draft” of end-of-high-school standards in those two subjects, and asked for comments within thirty days. Further revisions will doubtless follow. Meanwhile, drafting teams are beginning to “backward map” the standards down through earlier grades.

Nobody knows how all of this will turn out, but we think it’s worth pitching in to try to help make it turn out well—while reserving final judgment until we see the final product. Meanwhile, dozens of questions deserve to be asked and eventually need to be answered. Here’s a smattering:

Conceptually...Will the Common Standards in reading/writing/listening/speaking and math be as good as America needs in the

21st Century? Who says so and how do we know? How will we ensure their rigor and relevance over the long term? What about the assessments needed to give them traction? Where will those come from? Who will set their passing scores?

Logistically...What challenges (and landmines) lie ahead? What are the benefits and risks to states that adopt these standards? How will they know if these are better (or worse) than those they're using today? What is the intersection between "common core" standards and the next iteration of NCLB? Will such standards make it more or less likely that students will be deemed "proficient" and their schools will make adequate progress? What are the costs? What is the timeline for states' adoption? How many will adopt? Will anyone monitor their implementation for fidelity and if so, who?

Organizationally...What entity will "own" these standards and the tests that must follow? Who will be responsible for revising them? Correcting errors? Ensuring that assessment results are reported in timely fashion? Will comparative state data be seen as a blessing or a curse? How, if at all, will this venture relate to the federal government? Who will pay for it? What ancillary materials and tools for teachers should state governments or others provide? Will standards and tests in other subjects follow?

We're optimistic that these questions can be satisfactorily answered, tomorrow if not today. For years, we've favored movement toward national standards and tests. That preference has strengthened as we've learned more about the downside of fifty separate state standards and tests. But we also know that the stakes are exceptionally high. And we *really* want to see these standards done right. That means we must do our part, too.

And that, in turn, means helping to inform two parts of this debate. First, states need to know if these Common Core standards are any good—and better than their own. Second, as those developing the common standards amplify and refine their work—and as state officials work on their own standards, which they've been revising for years—they often look to existing national and

MEET THE EXPERTS

➤ **Sheila Byrd Carmichael** served as lead reviewer for English language arts. Ms. Carmichael is an education consultant based in Washington, D.C. who has taught English in the District of Columbia Public Schools and in Italy and Japan. She was the founding director of the American Diploma Project and is the former deputy executive director of the California Academic Standards Commission. She is the co-author of *Advanced Placement and International Baccalaureate: Do They Deserve Gold Star Status?* (Thomas B. Fordham Institute, 2007) and of *Why We're Behind: What Top Nations Teach Their Students But We Don't* (Common Core, 2008).

➤ **Carol Jago** served as external reviewer for English language arts. Ms. Jago is the director of the Reading and Literature Project at UCLA and incoming president of the National Council of Teachers of English (NCTE). She has served on the California English Language Arts Framework Committee, the National Assessment Governing Board (and its reading and writing framework planning committees), PISA steering committees, and the Common Core State Standards Initiative feedback group. Among her many awards from local, state, and national organizations is NCTE's 2006 Teacher of Excellence Award.

➤ **W. Stephen Wilson** served as lead reviewer for mathematics. Dr. Wilson is professor of mathematics at the Johns Hopkins University, where he has chaired the Department of Mathematics. In 2006, he was the advisor

for mathematics in the Office of Elementary and Secondary Education at the U.S. Department of Education. Dr. Wilson also helped revise Washington State's K-12 mathematics standards and evaluated textbooks for the state. He has participated in numerous projects on standards, curricula, and textbooks and co-authored *The State of State Math Standards* (Thomas B. Fordham Institute, 2005). More recently, he reviewed drafts of the Common Core mathematics standards for the National Governors Association and the Council of Chief State School Officers. Dr. Wilson received his Ph.D. in mathematics from M.I.T. in 1972. He has published over sixty research papers in mathematics, many in algebraic topology.

➤ **Richard Askey** served as external reviewer for mathematics. Dr. Askey is professor emeritus of mathematics at the University of Wisconsin-Madison. He is a fellow of the American Academy of Arts and Sciences and an honorary fellow of the Indian Academy of Sciences. He has had a Guggenheim fellowship and was vice president of the American Mathematical Society. His extensive writings on mathematics education include *Good Intentions are not Enough*, *Learning from Assessment*, and *Knowing and Teaching Elementary Mathematics*. He also served as reviewer of the National Research Council's report, *Adding It Up: Helping Children Learn Mathematics* and as lead content expert for the Indiana State Mathematics Standards. In 1999, he was elected to the National Academy of Sciences.

international tests and frameworks for guidance and models. “International benchmarking” is the new buzz word. So is “world class.” But which international and national frameworks and tests are worth benchmarking to?

OUR REVIEWERS AND A SNAPSHOT OF WHAT THEY FOUND

To help answer these questions, we enlisted four top-notch experts, individuals who not only possess deep content expertise in their respective fields, but who have also rolled up their sleeves to work with state officials and classroom teachers to ensure that state standards are crafted and implemented with integrity.

Sheila Byrd Carmichael served as lead reviewer for English language arts. She is an education consultant based in Washington, D.C., who has taught English in the District of Columbia Public Schools as well as Italy and Japan. She was the founding director of the American Diploma Project. **Carol Jago** served as external reviewer for English language arts. Ms. Jago is the Director of the Reading and Literature Project at UCLA and incoming President of the National Council of Teachers of English.

W. Stephen Wilson served as lead reviewer for mathematics. He is Professor of Mathematics at the Johns Hopkins University and has served as Senior Advisor for Mathematics at the U.S. Department of Education. **Richard Askey** served as external reviewer for mathematics. Dr. Askey is Professor Emeritus of Mathematics at the University of Wisconsin-Madison, Fellow of American Academy of Arts and Sciences, and in 1999, was elected to the National Academy of Sciences. (For more on our reviewers, see “Meet the Experts,” page 5.)

WHAT’S THEIR VERDICT?

- **PISA** • PISA strikes out. Neither in reading (literacy) nor in math does its content deserve better than a grade of “D.” This is not a promising benchmark for American K-12 education in these subjects.
- **NAEP** • NAEP fares better, with a “C” for math and “B” grades in reading and writing. But it ought to be better than it is.
- **TIMSS** • TIMSS does really well, earning an “A” in math. (Math and science are all that TIMSS touches.)
- **Common Core: Math** • The draft Common Core end-of-high-school standards in math, mid-September version, are better than PISA and NAEP, not as good as TIMSS. Our reviewers give this draft a “B” and offer suggestions for improving the final version.
- **Common Core: ELA** • The draft Common Core end-of-high-school standards in reading-writing-speaking-listening also earn a “B” from our reviewers—as well as much advice for strengthening and augmenting the final version.

THE ELA DILEMMA

That last point needs amplification, for it’s crucial to understand that drafters of the Common Core standards in reading, etc., faced a dilemma. If they tried to set standards for the entirety of English language arts, they would be inviting unwinnable battles over reading lists, authors, multiculturalism, and such. If, on the other hand, they confined themselves to the essential “skills” associated with reading/writing/speaking/listening, they would be open to complaints that “this isn’t really English” and “reading cannot be taught *sans* content.”

So they made a prudential judgment. The actual standards they set forth are limited to key skills but they carefully explain the types and levels of reading materials (including but not limited to literature) that they judge to be suited to those skills; they offer a few well-chosen illustrative passages; they underscore the interdependence of skills and content; and they state clearly and bluntly that these standards need to be accompanied by a rich, content-based curriculum. But they don’t supply that content themselves.

Someone must. This is really, really important, particularly for reading. In addition to our reviewers’ commentary on this point, consider these trenchant observations by University of Virginia psychologist Daniel Willingham, regarding the draft reading standards:

At first glance the 18 standards sound quite sensible: students should be able to determine what a text says, make inferences from it, discern the most important ideas, and so forth....The problem is that teachers and administrators are likely to read those 18 standards and to try to teach to them. But reading comprehension is not a “skill” that can be taught directly....The mainspring of comprehension is prior knowledge—the stuff readers already know that enables them to create understanding as they read....

What happens if the reader doesn’t have the prior knowledge the writer assumed she had? The reader will be confused and comprehension breaks down....If you take kids who score poorly on a reading test and ask them to read on a topic they know something about (baseball, say, or dinosaurs) all of a sudden their comprehension is terrific—better than kids who score well on reading tests but who don’t know a lot about baseball or dinosaurs.

In other words, kids who score well on reading tests are not really kids with good “reading skills.” The kids who score well on reading tests are ones who know a lot about the world—they have a lot of prior knowledge about a wide range of things—and so that whatever they are asked to read about on the test, they likely know something about it....

How do students get prior knowledge? It accumulates through years of exposure to newspapers, serious magazines, books, conversations with knowledgeable people. It should also come from a content-rich curriculum in school....

The new national standards actually say that....But the standards themselves don’t recommend that we ensure that students “have a strong content base” as a way to ensure that they are good readers! Instead, the standards document lists things that students ought to be able to do (summarize, find the main idea, etc.) that invite states, districts, and teachers to design curricula emphasizing practice in those skills.

The mistaken idea that reading is a skill—learn to crack the code, practice comprehension strategies and you can read anything—may be the single biggest factor holding back reading achievement in the country.

Much work awaits the drafters, at least in reading/writing/speaking/listening, both as they revise the end-of-high-school standards and as they backward-map these to earlier grades. And even more work awaits those who will deploy these standards as part of—please, just part of—their own expectations for schools and students.

We will, of course, have more to say on this topic when the revised versions come out. We naturally hope that the revision process will heed our reviewers’ comments, criticisms, and advice.

THE PATH AHEAD

The purpose of this report, then, is to assess the content, rigor, and clarity of the first public drafts of the CCSSI and—applying identical criteria—also to appraise these elements in America’s current de facto national math and reading/writing tests (NAEP) as well as two influential international testing regimes (PISA and TIMSS) that many look to for “benchmarking” purposes. Our subject experts have compared these test frameworks and standards using the same metrics. (See “Assumptions” on page 8 for a discussion of why testing frameworks may be considered standards.) Our goal is to help readers and users determine which of these documents ought to influence their thinking, their standards, their aspirations, and their tests. Is one of these national/international frameworks worthier of emulation than others? Are components of several stronger in different domains?

But this is an interim report. What we review here are the first public drafts released by the Common Core project. Made available on September 21, the public has been invited to provide feedback on them through October 21 via the CCSSI website. And that’s what we and our expert reviewers are doing. The common standards are still a work in progress;¹ hence, these reviews are not summative. That’s true of NAEP, PISA, and TIMSS, too. Some of them are also in the process of being updated or revised.

So we’ll be back in the spring of 2010 with a follow-up report. It will include new reviews of standards and tests from organizations frequently regarded as key players in determining college- and career readiness—the ACT and the College Board (both of which are partners in the Common Core initiative).

More specifically, the next report will round out our examination of math and English language arts standards. In math, we expect to add reviews of 1) the PISA 2009 Mathematics Frameworks (scheduled to be released in December 2009); 2) the TIMSS 2011 Mathematics Frameworks (just released in September 2009); 3) the ACT Testing Specifications for Mathematics; 4) the

1 The standards are being reviewed by an “expert Validation Committee” charged with evaluating the substance of the standards and their development process. The members of the Validation Committee can be found here: <http://www.corestandards.org/Files/CoreValidationCommitteeBios.pdf>. Members of the Standards Development Work Group (SDWG), along with additional experts, are in charge of the back-mapping. The members of the SDWG can be found here: <http://www.corestandards.org>.

College Board’s Standards for Mathematics and Statistics; and 5) revised versions of the Common Core standards for end-of-high-school mathematics as well as the “back-mapped” standards for earlier grades (assuming these are ready in time to review).

In English language arts/reading/writing etc., we expect the next report to review 1) the College Board’s Standards for English Language Arts and Writing; 2) the ACT Testing Specifications for English/Writing and Reading standards; 3) the PIRLS 2011 Assessment Framework (released in late summer 2009); and 4) revised versions of the Common Core end-of-high school standards in reading, listening, and speaking and listening as well as the back-mapped standards for the earlier grades (again, assuming these are ready).

Nor is that the end. The ELA/reading and mathematics reviews will be supplemented with reviews from the equally important core subject areas of science and history. Like many, we’re concerned that ELA and mathematics hog the academic spotlight and it’s critical that U.S. history and science also remain bulwarks of the K-12 curriculum, with solid standards and more. So, we’ll add four reviews in science and one in history. They include 1) the NAEP 2009 Science Frameworks; 2) the TIMSS 2009 Science Frameworks; 3) ACT Testing Specifications for Science; 4) the PISA 2009 Scientific Literacy Frameworks; and 5) the NAEP 2006 History Frameworks. (That this is both the only history review we have to offer and the latest version of NAEP speaks volumes about the neglect of this essential subject).

Amidst this flurry of national/international activity, Fordham will also commence another round of reviews of states’ current academic standards in English language arts, math, science, and history. (We last did this in 2005—and many states have since revised their standards.) Our purpose here is both to give state officials ongoing feedback regarding the standards that they’re actually using and to equip them and us with the capacity to make informed comparisons between state and national standards. We hope to release our next State of State Standards report in late 2010.

So, stay tuned. Like American education, we have miles still to travel on this twisty road.



This report was a collaborative product that benefitted from the expertise, skills, and hard work of many individuals. W. Stephen Wilson and Sheila Byrd Carmichael were the lead mathematics and English language arts reviewers (respectively) and authors of the reviews. In addition to their subject matter expertise, each is a consummate professional, willing to meet outrageous deadlines and to persist through multiple drafts. Carol Jago served as external reviewer for English language arts and Richard Askey for mathematics. Both were conscientious and thorough in their feedback. Thanks also to Andreas Schleicher, head of the Indicators and Analysis Division at OECD, for providing us with an advance copy of the PISA 2009 Reading Frameworks and for responding expeditiously to our other questions and requests. (We regret that he likely won’t welcome the reviewers’ conclusions.)

This report was made possible in part by grants from The Brookhill Foundation and the Bill and Melinda Gates Foundation. It was a joint effort with our sister organization, the Thomas B. Fordham Foundation, which also contributed financially. We appreciate the confidence, patience, and intellectual guidance provided by all of these funders. Quentin Suffren served as our adept copy editor. Alton Creative created the nifty design. At Fordham, interns Jack Byers and Janie Scull were quick to lend a helping hand. Amy Fagan and Laura Pohl capably handled dissemination, and Stafford Palmieri skillfully managed report production (in addition to other writing and editing tasks). Special thanks go to senior editor Marci Kanstoroom, who helped to shape the project along the way, as well as provided feedback and plenty of editorial assistance.

ORGANIZATION OF THE REPORT

The remainder of Part 1 first describes the assumptions that underpin our work, then the study methods. Part 2 describes how the standards were graded, including the content-specific criteria that our expert reviewers used to assess them and the grading metric employed to score them. Part 3 includes the standards reviews with math first (Common Core, NAEP, TIMSS, PISA) followed by English language arts (Common Core, NAEP, PISA). We close (part 4) with a takeaway. Finally, the appendix (part 5) includes an overview of the standards that have been reviewed.

Assumptions

■ CHESTER E. FINN, JR., AND AMBER M. WINKLER

This project, and those to follow, have operated under some key assumptions that merit discussion. They include 1) whether it is fair to judge all of the content-and-test frameworks included in this report as “standards”; 2) whether it’s appropriate to demand

that standards be “evidence based”; and 3) why the call for “fewer, clearer, higher standards” is more complicated than it seems. We discuss each below.

IS IT FAIR TO JUDGE CONTENT-AND-TEST FRAMEWORKS AS STANDARDS?

Most of the documents reviewed in this report do not claim to be content standards per se (with the exception of CCSS). They are “frameworks,” “guidance,” even “testing specifications.” But they stop short of representing themselves as standards. Is it legitimate, then, to appraise them as if they were standards?

Frankly, the lines between what standards claim to be, what they are, and how they’re used are often blurred. The PISA framework, for instance, has as its objective to test “not so much in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life.” Yet it goes on to “define the content that students need to acquire, the processes that need to be performed and the contexts in which knowledge and skills are applied.” In other words, though it initially claims not to incorporate content standards, PISA later offers up the very content it professes to exclude, which now serves as an “implicit” (if not very good) set of standards.

In this way, the PISA framework has the same intention and performs essentially the same function as academic standards, and can reasonably be evaluated by the same criteria. Further, a number of American educators, policy makers, and standard setters look to PISA as a benchmark for what they should require and/or expect of American schools. (We think it’s no coincidence that the head of PISA has just been appointed to the CCSS “validation committee,” along with his boss at the OECD.)

Much the same can be said for NAEP and TIMSS and the roles they play in American education generally and in standards-setting and benchmarking particularly. Since all these frameworks get used for these purposes, it is important to appraise their content. Hence, we use this opportunity to evaluate the NAEP, PISA, TIMSS, and Common Core reading/writing/communications and math frameworks as content (as well as skill) standards. Further, it’s long been observed that “what gets tested is what gets taught” so it’s more than reasonable to review assessment items (which we’ve done here when there were enough of them) and assessment frameworks in light of their likely impact at the classroom level.

That said, in the case of English language arts, reviewers confined their appraisals to the elements of ELA that the frameworks actually addressed. For instance, the Common Core standards are confined to reading, writing, and communications and do not attempt to encompass the full “content” of English language arts (e.g., literature and documents). Indeed, the drafters stipulate that their standards need to be accompanied by rich content-based curricula. Therefore, the draft Common Core standards are appraised as reading, writing, and communications standards, not ELA standards. The NAEP Reading and NAEP Writing Frameworks purport to be just what their names suggest—frameworks by which to design assessments of reading and writing performance—and are evaluated accordingly. Here, as in CCSS, it will be very important for state officials to understand those limitations and to appreciate what else may be needed to generate complete standards and curricula for English language arts.

IS IT APPROPRIATE TO DEMAND THAT STANDARDS BE “EVIDENCE BASED”?

Our second assumption is that expert judgment is a vital component of appraising academic standards. We respect the yearning for all standards to be “evidence-based” and the criteria used to assess standards in this report reflect that impulse. When real evidence is available, we ought to use it; ACT, for one, and researchers like William Schmidt at Michigan State have done much to expand the evidentiary base for standards development. Examining readiness for college via actual college work samples is also promising. But real evidence to date is quite limited. Perhaps that’s why CCSS drafters cast such a wide evidentiary net.

Yet merely finding and naming a country that has similar content in its standards or locating similar content in another test or framework (e.g., the ACT or American Diploma Project) is not compelling evidence. In fact, most of these determinations were originally derived from expert opinion, preference, or survey results. They were not an attempt to “validate” the standard per se, inasmuch as that means that the standard is necessary to accomplish a higher standard or end goal. In other words, much of the cited evidence in the common standards is suggestive, not dispositive. Frankly, we’re not that optimistic that all academic standards can or must be “validated,” partly because we don’t define a good education strictly in terms of college-and-career readiness, and partly because true validation studies are hard to do well. (We observe, too, that the recently-named “Validation Committee” is light on card-carrying mathematicians and literature experts.)

Moreover, strict adherence to “evidence” may belie old fashioned common sense. Certainly the focus on “college and career readiness” is well-intended and legitimate as far as it goes. Students must be able to deal with the academic and occupational challenges that lie ahead. But young Americans must also be equipped to contribute in our democracy as citizen-participants.

In her 1987 seminal text, *Democratic Education*, Amy Gutmann wrote: “The democratic interpretation of equal education opportunity requires that all educable children learn enough to participate effectively in the democratic process.” An education, then, is not only about preparation for what follows immediately after high school, namely college and/or job, but also about preparation for citizenship, community life, and cultural participation. If we define the purpose of education in narrow, utilitarian terms, and then build standards atop that definition, we could end up with a population that is highly skilled yet ignorant, self-absorbed, amoral, and uncultured.

WHY IS THE CALL FOR “FEWER, CLEARER, HIGHER STANDARDS” MORE COMPLICATED THAN IT SEEMS?

There’s been much talk about the need for “fewer, higher, and clearer” standards. At a glance, that’s a reasonable-sounding call to action. Among the failings of the “voluntary national standards” projects of the early 1990’s was their creators’ tendency to put into them everything but the kitchen sink, resulting in whopping fat volumes that nobody could realistically expect to use in actual schools. Yet “fewer” does not necessarily lead to clearer or higher; one might have a few standards that are vague and lack depth. Further, we’re inclined to believe that fewer, clearer, and higher standards work better in some disciplines than others. Math, for instance, is a single, cumulative discipline, quite unlike, say, English language arts, history, and science. A mathematics standard like “Compare and order whole numbers” is clear and explicit while a reading standard such as “Read and analyze both literary and non-literary texts” begs for clarification and explication. Perhaps it means that students need to evaluate the development of literary elements, or the effectiveness of rhetorical techniques, or the manipulation of stylistic devices? Perhaps they should describe the truth and/or validity of an argument or recognize and explain the presence of fallacious reasoning? Either way, it’s going to take more detail to transform that lofty but nebulous standard into an explicit one that’s actionable in the curriculum and the classroom.

Methods

■ AMBER M. WINKLER

Fordham has been reviewing state content standards in core subject areas (English language arts, math, science, history) for more than a decade. But as America moves in the direction of common, multi-state, or national standards, and as state officials (and “Common Core” drafters) seek to benchmark, model, or justify their work on the basis of extant national and international standards and tests, it was necessary to go farther, to determine, in effect, whether the models deserve to be emulated. So we replicated at the national/international level the methods we have previously used to evaluate state standards.

That meant relying once again on content experts. This time, we tapped four well-known mathematics and English language arts authorities, all of whom have worked with us in the past and have substantial experience working in the state standards arena (see “Meet the Experts” on page 5). Two served as lead reviewers; the other two judiciously responded to their work.²

We asked the reviewers, based on the best current thinking about their discipline—and a modest but growing evidence base around what students need to know and be able to do in order to be prepared for college, work, and democratic citizenship—to devise a set of content expectations for students at grades four, eight, and twelve (as well as age fifteen). These are the grade levels (and ages) at which NAEP, PISA, and TIMSS are administered. We encouraged them to consider the most important elements of content that standards should cover. In other words, we asked them to set forth a minimal set of crucial content expectations, not to dump into the sink every stray fork and teacup in their fields. Multiple drafts were exchanged and revised before the content criteria were finalized.

To facilitate apples-to-apples comparisons, we developed a common grading metric upon which all standards could be measured. In other words, reviewers would assess the various standards and tests according to content-specific criteria, but would need to render a bottom line grade based on a standardized scoring system. This meant that a “B” in mathematics would need to mean the same thing as a “B” in English language arts.

The final versions of the subject-specific criteria and common grading metric can be found next in part 2.

2 The lead reviewers benefited enormously from the advice of the outside reviewers (and feedback from Fordham and others). In the end, however, they take responsibility for the judgments and grades rendered.

The Grading

Math Content-Specific Criteria

■ W. STEPHEN WILSON with RICHARD ASKEY

This section lays out the core or minimal content material that students in fourth, eighth, and twelfth grade should know, as well as fifteen-year-old students (above and beyond eighth-grade expectations). It also includes an explanation of the proper treatment of data analysis, statistics, and probability in K-12 education (see “How much DASP do students need?” page 12).

FOURTH GRADE

Addition and subtraction

- 1 Students should know the single-digit addition facts and the corresponding subtraction facts.
- 2 They should be able to add and subtract whole numbers fluently and understand the algorithms they use.
- 3 They should understand the inverse nature of addition and subtraction.

Multiplication and division

- 1 Students should know the single-digit multiplication facts and be fluent with multiplying whole numbers.
- 2 They should understand the algorithms they use.
- 3 Students should be introduced to division and understand the inverse nature of multiplication and division.

Measurement

- 1 Students should understand some forms of measurement. In particular, they should know and understand the formulas for the area and perimeter of a rectangle and know how to measure distances to the nearest centimeter and half-inch using rulers.

Problem solving

- 1 Students should be able to use the above essential material to solve complex multi-step exercises and word problems.
- 2 In addition to the above core that should be mastered, they should be given an introduction to other material. Students should be introduced to fractions and decimals and be able to compare the magnitude of simple fractions and decimals.
- 3 Students should be able to use commutativity and associativity.
- 4 They should be able to do elementary estimations.
- 5 They need a basic vocabulary for geometry, including parallel and perpendicular lines.
- 6 They should be exposed to very elementary statistics and probability including reading and making basic graphs and tables.

EIGHTH GRADE

Arithmetic

- 1 Students should be fluent with the four arithmetic operations with rational numbers and decimals, without a calculator, and they should understand the operations.

Ratios

- 1 Students should understand and be able to use rates, ratios, proportions, and percentages.

Geometry and measurement

- 1 Students should have a good geometry vocabulary, know about angles associated with triangles and parallel lines, understand

similar and congruent triangles, and be able to compute areas, perimeters and volumes of various geometrical shapes, including circles.

- 2 They should know the Pythagorean Theorem and the coordinate system.

Problem solving

- 1 Students should be able to use the above essential material to solve complex multi-step exercises and word problems.
- 2 In addition to the essentials above, students should have an elementary knowledge of data analysis, statistics, and probability. (See below, “How much DASP do students need?”)
- 3 They should understand absolute value and be able to solve elementary linear equations.

FIFTEEN YEAR-OLDS

Fifteen year-olds should have mastered eighth-grade material and they should have the equivalent of a year of algebra as well. They should understand and be able to use roots, reciprocals, and powers. They should be able to solve equations and inequalities that are linear or involve the absolute value. Students should know about slope and forms of linear functions and how to graph them. Students should be fluent with the arithmetic of polynomials and rational expressions, including elementary factoring. Students should be able to use the above essential material to solve complex multi-step exercises and word problems. In addition to the essentials above, students should be introduced to quadratic equations and their solutions.

TWELFTH GRADE

Minimal twelfth grade expectations, in addition to the expectations for fifteen-year-olds, include a working knowledge of geometry containing constructions, definitions, and proofs of the major results in Euclidean geometry. Students should understand logarithmic and exponential functions as well as basic trigonometry and trigonometric functions. Students should also be able to completely analyze quadratic equations, inequalities, and functions. Students should know basic statistics and probability, particularly the counting arguments involving combinations and permutations. Students should be able to use this essential material to solve complex multi-step exercises and word problems.

HOW MUCH DASP DO STUDENTS NEED?

■ W. STEPHEN WILSON

Data Analysis, Statistics, and Probability (DASP), play a prominent role in many sets of standards. Certainly students should be able to read and construct various kinds of graphs and displays for data, compute elementary probabilities, and know the basic descriptors for data such as percentiles, mean, and median. In addition, they should be able to use combinations and permutations in their counting arguments for probability and have some concept of random variability. Many standards, however, go far beyond these basic needs. Justification for this is usually slim, frequently based on a vague notion that a good citizen should know this material. Presently, many standards include deep mathematical content that is far above the level of mathematics accessible to K-12 students. This results in superficial coverage that often lacks mathematical coherence. In addition, the “grain size” of DASP standards tends to be smaller than for other areas of math standards, i.e., more detail is devoted to them. This increases the number and percentage of standards dedicated to DASP and gives them more prominence than they deserve.

A common example, frequently observed in middle school standards, is a variant on “eye-balling” a line of best fit for a scatter-plot. More specifically, the eighth grade NAEP framework has this standard: “Visually choose the line that best fits given a scatterplot.” “Visually choosing” (i.e., eye-balling) is not a mathematical activity. Further, the concept of “least squares” is mentioned in the NAEP twelfth grade standards, though it is inappropriate to include it. That’s because most students do not have the requisite math background to see this for what it is: Finding the closest point on a plane in n -dimension Euclidean space to a point not on that plane using the Pythagorean Theorem to compute the distances. This is material from a college-level linear algebra or multivariable calculus course, courses that generally follow calculus. Another example is when students are exposed to the normal distribution, but its origin and connection to the binomial distribution—college-level material to be sure—is not covered, leading to superficial coverage of the former. There are many other examples where students lack mathematical understanding with regard to what is actually occurring in a DASP problem.

Further, recent research has shown that DASP does not prepare students for placement exams in college or for college level material,³ nor is it material requested by college mathematics teachers.⁴ Certainly students headed for scientific and technological careers will be required to take college-level courses in statistics and probability, as will anyone else who needs this material in college. In the end, though, knowledge of arithmetic, ratios and algebra will be much more important to *all* students than DASP content void of a conceptual base. After all, students should possess the mathematical foundation that allows them to understand what they are doing.

English Language Arts Content-Specific Criteria

■ SHEILA BYRD CARMICHAEL with CAROL JAGO

The following criteria contain examples in many places to clarify the type (and level) of concepts and skills that are expected in quality standards at the designated grade spans. The criteria and examples for each grade span should be understood to include the criteria and examples for the grade spans that precede them.⁵

GRADES K-4

Reading

- 1 The standards delineate explicit and systematic expectations in phonemic awareness, phonics, fluency, and comprehension skills.
- 2 The standards address systematic vocabulary development (e.g., basic prefixes and suffixes; common synonyms, antonyms, and compound words; multiple meaning words; dictionary use).
- 3 The standards outline specific expectations for reading and analyzing literary and non-literary texts (e.g., recognizing and interpreting genres; structures; literary elements; and stylistic devices).
- 4 The standards reflect the importance of reading grade-appropriate works of outstanding American literature that reflect our common heritage.
- 5 The standards describe the amount, quality, and complexity of both literary and non-literary texts to be studied through the use of lists (authors and/or titles), sample passages, and/or commentary.

Writing

- 6 The standards delineate expectations for writing that address the characteristics and quality of writing products appropriate to the grade level (e.g., organization of ideas and focus; introduction, body and conclusion; elements of a paragraph; evaluation and revision skills).
- 7 The standards require students to recognize, explain, and produce writing that reflects the defining characteristics of various grade appropriate writing genres (e.g., narration; exposition).
- 8 The standards describe or reference the use of specific criteria for evaluating writing (e.g., logically organized and detailed genre- or prompt-specific rubrics) that include examples regarding the quality of writing expected.

Listening and Speaking

- 9 The standards clearly address active listening and effective speaking skills (e.g., summarizing information presented orally; asking and answering relevant questions).
- 10 The standards address the ability to make formal oral presentations (e.g., recitation; story retelling; and sequencing).
- 11 The standards describe or reference the use of specific criteria for evaluating oral presentations (e.g., content; organization; and presentation style).
- 12 The standards include specific expectations for participation in group discussions (e.g., turn-taking; applying agreed-upon rules for decision making).

3 Achieve, Inc. *Aligned Expectations: A Closer Look at College Admissions and Placement Tests*. Washington, D.C.: Author. http://www.achieve.org/files/Admissions_and_Placement_FINAL2.pdf.

4 ACT. 2007. *ACT National Curriculum Survey 2005-2006*. Iowa City, Iowa: Author, 41-43. <http://www.act.org/research/policymakers/pdf/NationalCurriculumSurvey2006.pdf>.

5 In other words, standards for fifth- to eighth-graders should be built upon the material already covered by the criteria and examples for students in grades K-4. Further, standards for grades nine through twelve should incorporate both previous age groups.

Oral and Written Language Conventions

- 13 The standards specify expectations for the correct use of standard English, describing a grade-appropriate facility with the parts of speech, sentence structure, usage, and mechanics appropriate to the grade level (e.g., nouns, verbs, adjectives, adverbs, conjunctions, prepositions, and nominative/objective/interrogative pronouns; sentence types; complete/incomplete sentences; subject/verb (S/V) agreement; initial, internal, and ending punctuation; and basic spelling rules, such with as plurals, contractions, and inflections).

Research

- 14 The standards require students to learn the research process, outlining specific expectations for the essential components of the process (e.g., identifying or finalizing a research question; locating information; evaluating and compiling information; presenting findings; and acknowledging sources using a standard format).

Media

- 15 The standards require students to analyze and evaluate information presented in multimedia formats (e.g., the effect of various visual and aural techniques; how information presented in print is different from that which is presented through the use of multimedia).
- 16 The standards require that students learn about multimedia techniques for presenting information.

GRADES 5-8**Reading**

- 1 The standards address vocabulary development (e.g., knowledge of roots and affixes; connotation and denotation; figurative language; use of the dictionary for clarifying multiple meanings, etymology, and pronunciation).
- 2 The standards specify strategies/skills for reading and analyzing both literary and non-literary texts (e.g., analysis of genres, structures, literary elements, rhetorical techniques, and stylistic devices; strategies for comprehension and interpretation).
- 3 The standards reflect the importance of reading grade-appropriate works of outstanding American literature that reflect our common heritage.
- 4 The standards describe the amount, quality, and complexity of both literary and non-literary texts to be studied through the use of lists (authors and/or titles), sample passages, and/or commentary.

Writing

- 5 The standards delineate expectations for writing that address the characteristics and quality of writing products appropriate to the grade level (increasingly sophisticated understanding of audience and purpose; clear organization and consistent focus; development of ideas through multi-paragraph essays; use of transitions; reflective peer review and revision processes).
- 6 The standards require students to interpret and produce writing that reflects the defining characteristics of various grade-appropriate writing genres (e.g., argument).
- 7 The standards describe or reference the use of specific criteria for evaluating writing (e.g., logically organized and detailed genre- or prompt-specific rubrics) that include examples regarding the quality of writing expected.

Listening and Speaking

- 8 The standards clearly address active listening and effective speaking skills (e.g., give, restate, and execute multi-step directions; convey ideas orally and interpret spoken ideas; make inferences from spoken information; ask and answer clarifying questions).
- 9 The standards address the ability to make formal oral presentations (e.g., recitation; informative and persuasive presentations that offer supporting details and evidence; address anticipated counterclaims and include a call to action when appropriate).
- 10 The standards describe or reference the use of detailed criteria for evaluating formal oral presentations.
- 11 The standards include specific expectations for participation in group discussions (e.g., designation of roles; eliciting and considering suggestions).

Oral and Written Language Conventions

- 12 The standards specify expectations for the correct use of standard English, describing a grade-appropriate facility with the

parts of speech, sentence structure, usage, and mechanics appropriate to the grade level (e.g., parts of the verb; interjections, possessive/demonstrative/relative/indefinite pronouns; tenses; analysis of sentence structure; types of phrases and clauses; fragments and run-on sentences; facility with mechanics grounded in understanding of sentence structure).

Research

- 13 The standards require students to employ the research process, outlining specific expectations for the essential components of the process (e.g., identifying and refining a research question; locating information; evaluating the quality of information/sources; selecting information that supports a thesis; presenting findings; citing sources correctly using standard guidelines; and avoiding plagiarism).

Media

- 14 The standards require students to analyze and evaluate information presented in multimedia formats (e.g., how information presented in print is different from that which is presented through the use of multimedia; noting what is conveyed through the use of various visual and aural techniques, such as bias and propaganda).
- 15 The standards require that students know how to use multimedia techniques to present information.

GRADES 9-12

Reading

- 1 The standards address vocabulary development and skills for building content area vocabulary (e.g., applying knowledge of roots and affixes to help determine meanings of words; tracing etymology; determining shades of meaning).
- 2 The standards describe specific expectations for reading and analyzing both literary and non-literary texts (e.g., analyzing the clarity of structures, the development of literary elements, the effectiveness of rhetorical techniques, and the manipulation of stylistic devices; describing the truth and/or validity of an argument; recognizing and explaining the presence of fallacious reasoning).
- 3 The standards reflect the importance of reading grade-appropriate works of outstanding American literature that reflect our common literary heritage.
- 4 The standards describe the amount, quality, and complexity of both literary and non-literary texts to be studied through the use of lists (authors and/or titles), sample passages, and/or commentary.

Writing

- 5 The standards delineate expectations for writing that address the characteristics and quality of writing products appropriate to the grade level (e.g., strong organization and development of ideas; facility with selection and blending of genres appropriate to audience and purpose; the use of sophisticated transitions, active rather than passive voice, and other stylistic elements for rhetorical effect).
- 6 The standards require students to analyze and produce writing that reflects the defining characteristics of various grade-appropriate writing genres (e.g., persuasion).
- 7 The standards describe or reference the use of specific criteria for evaluating writing (e.g., logically organized and detailed genre- or prompt-specific rubrics) that include examples regarding the quality of writing expected.

Listening and Speaking

- 8 The standards clearly address active listening and effective speaking skills (e.g., interpret complex information and ideas presented orally and convey complex information or ideas orally).
- 9 The standards address the ability to make formal oral presentations (recitation; complex informative or persuasive oral presentations that require a logical structure, well-chosen supporting evidence/details, skillful rhetorical techniques, and a strong presentation style).
- 10 The standards describe or reference the use of detailed criteria for evaluating formal oral presentations.
- 11 The standards include specific expectations for participation in group discussions (e.g., tolerating ambiguity; building on the ideas of others; reaching consensus).

Oral and Written Language Conventions

- 12 The standards specify expectations for the correct use of standard English, describing a grade-appropriate facility with the parts of speech, sentence structure, usage, and mechanics (e.g., demonstrate control of sentence structure, usage, and mechanics).

Research

- 13 The standards require students to conduct the research process, outlining specific expectations for the essential components of the process (e.g., identifying and refining a research question; locating information; evaluating the quality of information/sources; selecting information that supports a thesis; excluding extraneous information; presenting findings in a format appropriate for the audience and purpose; citing sources correctly in a standard format; and avoiding plagiarism).

Media

- 14 The standards require students to analyze and evaluate information presented in multimedia formats (e.g., noting instances of manipulation, bias, propaganda, and potential fallacies).
- 15 The standards require that students use multimedia techniques to present information when possible.

Common Grading Metric

■ JOINTLY DEVELOPED BY FORDHAM AND THE REVIEWERS

Reviewers compared each set of standards to subject-specific content expectations (see pages 11–16).⁶ Based on that comparison, they assigned the standards two scores, one for “content and rigor,” the other for “clarity and specificity.” “Content and rigor” is scored on a 0-7-point scale while “clarity and specificity” is scored on a 0-3-point scale, as follows.

CONTENT AND RIGOR

7: Standards meet all of the following criteria:

- Standards are top-notch in terms of the content chosen. The coverage of the subject is suitable, good decisions have been made about what topics to include, and nothing of importance has been overlooked. (No more than 5 percent of the content outlined in the subject-specific content expectations is missing.)
- Not only is the appropriate content covered by the standards, but it is covered *well* (i.e., in a high quality manner).
- Good decisions have also been made about what content should be left out. Excellent standards do not include much superfluous material. (No more than 5 percent of the content in the standards is unnecessary.)
- Standards distinguish between more important and less important content and skills either directly (i.e., by articulating which are more or less important) OR via the number of standards dedicated to particular content and skills (i.e., more important content/skills have more standards while less important content/skills have fewer standards). The standards do not overemphasize topics of little importance or underemphasize topics of great importance.
- The level of rigor is appropriate for the targeted grade level(s). Students are expected to learn the content and skills in a sensible order and an appropriately increasing level of difficulty. The standards, taken as a whole, define a core literacy for all students in the subject under review; at the same time, the standards that run through grade 12 are sufficiently challenging to ensure that students who achieve proficiency by the final year of high school will be ready for college or work and citizenship.
- The standards do not overemphasize the importance of students’ life experiences or “real world” problems. They do not embrace fads, suggest political bias, or teach moral dogma. They do not imply that all interpretations are equally valid (regardless of logic or the adequacy of supporting evidence). The standards also avoid other major subject-specific problems identified by the reviewers. While the standards are not perfect, any defects are marginal.

6 See the “Methods” section for a discussion of why we use “standards” to describe all of the documents reviewed in this report.

6: Standards fall short in one or more of the following ways:

- ❑ Some crucial content (as specified in the subject-specific content expectations) is missing (at least 5 percent and up to 20 percent).
- ❑ The content is covered satisfactorily but not in a high quality manner.
- ❑ Some of the content in the standards is unnecessary (at least 5 percent and up to 20 percent).
- ❑ Standards do not fully distinguish between more and less important content and skills (i.e., importance is neither expressly articulated nor conveyed via the number of standards dedicated to particular topics). In other words, the standards overemphasize no more than one or two topics of little importance or underemphasize no more than one or two topics of great importance.
- ❑ Standards at particular grade levels are *not quite* as rigorous as they could be, or are *too* rigorous (i.e., expectations are slightly too high or too low).
- ❑ There are minor problems or shortcomings (e.g., one or more of the problems listed in the last paragraph under score 7 affects the standards in a small way, or there are other minor subject-specific problems).

5: Standards fall short in one or more of the following ways:

- ❑ Some crucial content is missing (at least 20 percent and up to 35 percent).
- ❑ While most of the appropriate content is covered by the standards, the content is nonetheless covered in a manner that is not satisfactory (i.e., the standards cover the right material but do not cover that material robustly, thus, the material is shortchanged in some way).
- ❑ Some of the content in the standards is unnecessary (at least 20 percent and up to 35 percent).
- ❑ Standards do not distinguish between more and less important content and skills (i.e., importance is not articulated or conveyed in any way). The standards often overemphasize topics of little importance or underemphasize topics of great importance.
- ❑ Standards generally need to be more or less rigorous than they are at certain grade levels (i.e., expectations are too high or too low).
- ❑ There may be an important shortcoming (perhaps one of the problems listed in the last paragraph of score 7, or there are other subject-specific problems).

4: Standards fall short in one or more of the following ways:

- ❑ At least 35 percent and up to 50 percent of crucial content is missing.
- ❑ Some of the content in the standards is unnecessary (at least 35 percent, and up to 50 percent).
- ❑ There may be a few critical shortcomings (as listed above) although the standards contain no serious errors.

3: Standards fall short in one or more of the following ways:

- ❑ At least 50 percent and up to 65 percent of crucial content is missing.
- ❑ At least 50 percent and up to 65 percent of the content in the standards is unnecessary.
- ❑ There are serious problems or shortcomings or errors in the standards, although (the standards have some redeeming qualities and there is some evidence of rigor).

2: Standards fall short in one or more of the following ways:

- ❑ At least 65 and up to 80 percent of crucial content is missing.
- ❑ At least 65 percent and up to 80 percent of the content in the standards is unnecessary.
- ❑ There may be several serious problems, shortcomings, or errors (as listed above).

1: Standards fall short in one or more of the following ways:

- ▣ At least 80 percent of crucial content is missing.
- ▣ At least 80 percent of the content in the standards is unnecessary.
- ▣ There are numerous problems, shortcomings, or errors (as listed above).

0: Standards fall short in one or more of the following ways:

- ▣ The content of the standards does not address or barely addresses the subject-specific content expectations.
- ▣ The content is poorly chosen and fails to provide the level of rigor appropriate for the targeted grade level(s).
- ▣ Content is full of problems, shortcomings, and errors (as listed above).

CLARITY AND SPECIFICITY**3: Standards are coherent, clear, and well organized.**

The scope and sequence of the material is apparent and sensible. They provide solid guidance to users (students, teachers, curriculum directors, test developers, textbook writers, etc.) about the content knowledge and skills required to do well on the exam. The right level of detail is provided. The document(s) are written in prose that the general public can understand and are mostly free from jargon. The standards describe things that are measurable (i.e., can lead to observable, comparable results across students and schools). The standards as a whole clearly illustrate the growth expected through the grades.

2: The standards are somewhat lacking in coherence, clarity, or organization.

The scope and sequence of the material is not completely apparent or sensible. The standards do not quite provide a complete guide to users as to the content knowledge and skills required to do well on the exam (i.e., as a guide for users, there are shortcomings that were not already addressed by the content and rigor score). The standards provide insufficient detail. The prose is generally comprehensible but there is some jargon and some vague or unclear language. Some standards are not measurable.

1: The standards are somewhat coherent, clear, and organized.

They offer limited guidance to users (students, teachers, curriculum directors, textbook writers, etc.) about the content knowledge and skills required to do well on the exam, but there are significant shortcomings (as a guide for users) that were not already addressed by the content and rigor score. The standards are seriously lacking in detail, and much of their language is vague enough to leave unclear what is being asked of students and teachers.

0: The standards are incoherent and/or disorganized.

They are not helpful to users. The standards are sorely lacking in detail. Scope and sequence is a mystery.

FINAL GRADES

A final grade for each set of standards is calculated by adding the “content and rigor” score to the “clarity and specificity” score.

A = 9 or 10 points

B = 7 or 8 points

C = 5 or 6 points

D = 3 or 4 points

F = 0, 1, or 2 points

Reviews: Math

■ W. STEPHEN WILSON

- 1 Common Core State Standards for Mathematics
- 2 National Assessment of Educational Progress (NAEP) Mathematics
- 3 Trends in International Mathematics and Science Study (TIMSS) Mathematics
- 4 Programme for International Student Assessment (PISA) Mathematics

Common Core State Standards for Mathematics

DOCUMENTS REVIEWED

Common Core State Standards Initiative. 2009. *College and career readiness standards for mathematics: Draft for review and comment*. Washington, D.C.: National Governors Association Center and Council of Chief State School Officers. September 21. <http://www.corestandards.org/Standards/index.htm>

Introduction

The Common Core State Standards Initiative (CCSSI) is jointly organized by the National Governors Association and the Council of Chief State School Officers. As of September 2009, fifty-one states and territories are participating in the effort (Texas and Alaska are the exceptions). Though the mathematics standards reviewed herein are not the final documents, they are the first to be released to the public for general review and comment. The standards are designed to “define the knowledge and skills students should have to succeed in entry-level, credit-bearing, academic college course and in workforce training programs.” As such, they are “end of high school,” or, more accurately, “college readiness” standards.

The standards document is twenty-three pages, but the actual standards are broken down into ten content areas (such as number, equations, statistics), each neatly contained on one page and, on average, having five explicit “core skills” (i.e., standards, usually with some explanatory notes). With each content area there is a brief listing of “core concepts” that give an overview and a few paragraphs of commentary explaining what a “coherent understanding” of the area requires.

Review

This review has several parts. The standards’ content is compared to what should be there (see “Math Content-Specific Criteria,” page 11) and divided into three sections: “content covered,” “content missing,” and “content problems,” followed by a conclusion about content and rigor. Clarity and specificity are then examined, followed by a final summary and grade.

CONTENT COVERED

End of high school standards should thoroughly cover the arithmetic of rational numbers, including decimals, and also cover rates, ratios, proportions and percentages, as well as problem solving using these skills. Under the content heading “functions,” rates, ratios, and proportions are well covered. Percentages are covered under the heading “number.” As for the arithmetic of rational numbers, again under the heading “number,” we have:

- Know when and how to use standard algorithms, and perform them flexibly, accurately and efficiently.
- Solve multi-step problems involving fractions and percentages.

The phrase “standard algorithms” is generally applied to the four arithmetic operations with whole numbers. If the authors meant this to cover operations with rational numbers and decimals, they should make it explicit. Decimals are barely mentioned (once to compare and make sense of magnitude) and fractions are mentioned in that context and in the above standard.

Measurement of lengths, areas, etc., is nicely covered under the heading of “shape” with the explicit standard “solve problems involving measurements.” There is a paragraph giving more detail about this standard.

Geometry is also covered under the heading “shape.” Angles, parallel lines, and properties of triangles are covered. The Pythagorean Theorem also appears under shape but coordinates are given their own heading.

Absolute value, slope, solving linear equations and the arithmetic of polynomials, including factoring, are all included in the standards. Exponentials, basic trigonometry, and quadratic equations are there as well. Basic probability and statistics are present, including combinations and permutations.

CONTENT MISSING

The primary missing content is *explicit* reference to the arithmetic of rational numbers. The intent might have been to include this under standard algorithms (noted above) but it is not explicit. The only place logarithms appear is when students are asked to graph them using technology. This is inadequate coverage. Finally, the only explicit mention of the Pythagorean Theorem in the core skills is that it should be “used” to solve problems. The Pythagorean Theorem and its converse should also be proven.

CONTENT PROBLEMS

Linear trend lines creep into the standards in two places, but the necessary mathematical prerequisites for this are not there.

Content and Rigor Conclusion

The lack of explicit reference to the arithmetic of rational numbers is a serious oversight and falls under the heading “crucial content is missing.” In this case, the score of six is appropriate (i.e., between 5 percent and 20 percent of the content missing). The arithmetic of rational numbers certainly is more than 5 percent of the necessary content.

Although there are only fifty explicit core skill standards, these “standards do not fully distinguish between more important and less important content and skills” (see “Common Grading Metric,” page 16). For example, of the fifty standards, twelve of them are about probability and statistics. These standards are acceptable, but it’s highly unlikely that they will show up on a college placement test and students who do *not* go to college will not have jobs that require the use of this mathematics. The problem with these standards is that they occupy 24 percent of the core skills. They do this by being very detailed. The six standards under probability, for instance, could easily be combined into one or two standards, whereas many of the individual standards under other, more important, headings (e.g., “number,” “expressions,” “equations,” or “functions”) could easily be made more explicit by expanding them to several standards. For example, “manipulate simple expressions” requires a long explanatory paragraph as it is and could—and possibly should—be expanded to several standards.

The bottom line is that readers cannot tell which standards should be given more time than others. As indicated, the standards “do not fully distinguish between more important and less important content and skills.” The score for content and rigor is five.

Clarity and Specificity

The organizational structure of the Common Core standards is nicely done. The standards are coherent, clear, and well organized, and thus deserve a score of three out of three for clarity and specificity.

Summary and Grade

The format of the Common Core standards is well done. It breaks the standards up into ten content areas—a significant variation on the usual content areas that most extant state standards use, but it works well.

Common Core has kept its standards simple and clear. They are easy to understand and very few need to be rewritten or deleted.

The standards are not explicit enough about the arithmetic of rational numbers, but the worst feature of these draft standards is one shared with most standards: They do not set priorities. As a result, many standards that will contribute little (perhaps nothing) to college readiness are given equal status with standards that are essential foundations.

CONTENT	CLARITY	TOTAL	GRADE
5	3	8	B

National Assessment of Educational Progress (NAEP) Mathematics

DOCUMENTS REVIEWED

National Assessment Governing Board. 2008. Mathematics framework for the 2009 National Assessment of Educational Progress. Washington DC: US Department of Education. <http://www.nagb.org/publications/frameworks/math-framework09.pdf>

Introduction

The National Assessment of Educational Progress (NAEP) is known as “the nation’s report card.” NAEP mathematics exams are administered at grades four, eight, and twelve. The Mathematics Framework is seventy-four pages long and includes an overview of the exam, a description of the test design, sample exam items, and item formats, as well as a detailed list of content items—in essence, standards—for grades four, eight, and twelve. The standards are organized by strand (such as Measurement and Geometry), with grade levels placed side-by-side to show how they should develop from fourth to eighth to twelfth grade. Breakdowns are given of the percentage of exam items for each strand. Sample exam items appear throughout the framework to illustrate some of the standards and clarify aspects of the test design.

Review

The NAEP mathematics framework is evaluated on two dimensions: “content and rigor” and “clarity and specificity.” This review addresses these dimensions in several sections below. First, the framework’s content is compared with the content that should be covered by mathematics tests for fourth-, eighth- and twelfth- graders. (See “Math Content-Specific Criteria,” page 11.) Separate sections describe the content covered by the NAEP framework for each grade level, as well as what is missing from and problematic about it. The next section sums up the content and rigor of the NAEP framework across grade levels. Then the review appraises the clarity and specificity of the NAEP framework as a whole, and the last section offers an overall summary and final grade.

In brief, most of the important content that we would expect students to learn is covered by the NAEP mathematics standards. But since the framework includes nearly 300 standards, it is impossible to distinguish the important standards from the less important ones. (For high school algebra, for instance, the heart of algebra—representing the abstract with symbols—is inappropriately deemphasized, while equal status is given to several non-symbolic approaches.) In addition, some important content is missing, and for other content that is included, inadequate detail is provided.

A majority of the standards also have minor or major problems with clarity—meaning they are awkward, verbose or imprecise, or they contain meaningless phrases or verbs that do not make clear mathematical sense.

Fourth Grade

CONTENT COVERED

Whole number arithmetic (and problem solving using whole number arithmetic) should make up the bulk of learning in mathematics in grades one through four. In the NAEP mathematics framework, there are standards that cover whole number arithmetic and its application to problem solving.

The measurement of lengths in inches and meters, and the computation of perimeters and areas of rectangles, are also covered, as is problem solving.

Additionally, simple fraction and decimal comparison, including adding fractions with like denominators, is covered. Commutativity, associativity, and distributivity are presumably covered under the standard “Apply basic properties of operations,” and the inverse nature of addition and subtraction, and of multiplication and division, may also be covered by this standard, though there is no specific reference to these properties.

Estimation is covered and the basic vocabulary for geometry is introduced. Standards for elementary data analysis, statistics, and probability are presented shortly thereafter.

CONTENT MISSING

There is no mention of single-digit addition and multiplication facts and the corresponding subtraction and division facts. While sample problems are given that use some of these facts, students need not know the more complicated single-digit facts in order to solve them. There is no explicit reference to the inverse nature of addition and subtraction or multiplication and division. Also, there is no mention of how precise measurements should be.

Eighth Grade

CONTENT COVERED

Mathematics in grades five through eight should thoroughly cover the arithmetic of rational numbers, including decimals, and also cover rates, ratios, proportions, and percentages. These represent the core content on which students should spend the most time. Arithmetic is covered by the NAEP framework under the standard that says, “Perform computations with rational numbers”; there are also a couple of supporting standards. Other standards cover rates, ratios, proportions, and percentages. Similar triangles, perimeters, volumes, circles, applications of the Pythagorean Theorem, and an introduction to the coordinate system are all present in the geometry section. Problem solving is included with this geometric content.

Additionally, basic data analysis, statistics, and probability are well covered. Absolute value and basic linear equations are also included.

CONTENT MISSING

Ratios are covered inadequately by the vague standard “Use ratios to describe problem situations.” Problem solving with ratios is altogether missing, and there is no mention of the angles associated with a line cutting across parallel lines (e.g., vertical, corresponding, alternate interior/exterior angles, etc.).

CONTENT PROBLEMS

It is not unusual for frameworks or standards documents to include process or mathematical reasoning standards—standards that are not about math content, but about how to think mathematically. NAEP has such standards at the end of each content strand and they are often completely subjective. Nearly anything could be used to test them. For example, one standard reads, “Explain or justify a mathematical concept or relationship.” This nebulous admonition spans all of mathematics from kindergarten to the cutting edge of research.

There are also some mathematical misunderstandings, such as this one:

- | Interpret “=” as an equivalence between two expressions and use this interpretation to solve problems.

The language of this standard suggests that there are a variety of possible meanings for “=.” Yet the “=” sign is not up for interpretation—it actually means something.

Some of the many standards listed under data analysis, statistics, and probability (DASP) are descriptive or qualitative standards, not mathematical standards. (See “How much DASP do students need?” on page 12 for more about non-mathematical content in DASP standards.)

Consider this first example:

- | Given a graph or a set of data, determine whether information is represented effectively and appropriately.

This is not really a mathematics standard, as the answer is based on a value judgment rather than any mathematical foundation. Another standard reads as follows:

| Visually choose the line that best fits given a scatterplot and informally explain the meaning of the line.

This is not mathematics, and a line chosen to fit a scatterplot in this manner is mathematically meaningless. Nor it is clear what is meant by “informally explain.”

Twelfth Grade

CONTENT COVERED

Most of the content covered by the grade twelve framework breaks up nicely into geometry and algebra. Proofs of traditional Euclidean theorems are generally included along with definitions and the necessary structure of geometry in the geometry section.

The core material requires a thorough understanding of linear and quadratic equations and functions. Both linear and quadratic equations come in multiple forms, and being able to algebraically (i.e., symbolically) translate between the various forms is important.

Most of the linear material is covered by the eighth-grade standards, but there is some new material in this standard:

| Create and translate between different representations of algebraic expressions, equations, and inequalities (e.g., linear, quadratic, exponential, or trigonometric) using symbols, graphs, tables, diagrams, or written descriptions.

This would also seem to cover much of quadratics. Eight standards in all mention quadratics, including standards covering inequalities. An important part of the study of quadratic functions is understanding that the graph is symmetrical and learning to locate the maximum or minimum of the function. This opens up a whole new world of problems that can be solved and requires solid technical work that allows a student to transform the standard form of a quadratic function into its vertex form. The closest standard to this (albeit one with inadequate detail) reads as follows:

| Recognize and analyze the general forms of linear, quadratic, rational, exponential, or trigonometric functions.

Roots and exponents are already covered in the eighth-grade framework. Logarithms, exponentials, and basic trigonometric functions are all covered. Probability and statistics, including combinations and permutations, are more than adequately covered.

CONTENT MISSING

Among the content missing are geometric constructions. Explicit reference to the various forms of linear and quadratic equations, and translations between them, are also missing. Despite there being 120 standards for twelfth-grade mathematics (eight involving quadratics), significant details are missing, such as completing the square, the symmetry of the graph, and finding the maximum or minimum.

CONTENT PROBLEMS

The power of algebra is in its abstraction: the ability to reduce problems to symbols and solve them by manipulating the symbols. There are twenty-seven standards in the algebra strand for twelfth grade; however, algebra as abstraction, with the use of symbols, is not the emphasis. Consider again this twelfth-grade standard:

| Create and translate between different representations of algebraic expressions, equations, and inequalities (e.g., linear, quadratic, exponential, or trigonometric) using symbols, graphs, tables, diagrams, or written descriptions.

The objects under discussion are symbolic: algebraic expressions, equations, and inequalities. Yet the symbolic approach to algebra is only one of five approaches (along with graphs, tables, etc). This de-emphasis of the symbolic misrepresents the nature of algebra.

Content and Rigor Conclusion

The content covered by the NAEP standards at the fourth-, eighth-, and twelfth-grade levels would normally earn the standards a score of six; coverage is good, though there are some missing items. There are significant flaws in both the eighth-grade and

twelfth-grade standards, but the primary problem with the NAEP framework is its failure to set priorities. With nearly 300 standard-style items, all presented as equal, there is no way for the reader or user to determine which are the most important. Minor standards appear as equal to major ones. This is particularly evident in the algebra strand, where the words and phrases *tables*, *diagrams*, *graphs*, and *written descriptions* appear fifty-seven times while the more important focus for algebra, the *symbolic*, only appears thirteen times. Secondary content overshadows primary content simply because the secondary content appears so much more often. According to the “Common Grading Metric” (see page 16), a set of standards cannot score above a five for content if they “do not distinguish between more important and less important content and skills.” Consequently, a score of five is indicated.

Clarity and Specificity

The structure of the NAEP framework is acceptable. It shows the progression of content from fourth through twelfth grade. It is organized by content strand. However, it has a number of problems with clarity and specificity, most pertaining to non-mathematical verbs and unclear language.

NON-MATHEMATICAL VERBS

The NAEP framework is expressly designed to prepare students for the NAEP math exams, yet words like *identify*, *interpret*, *represent*, and *recognize* occur well over 100 times. These are not words that would be used in a math problem; indeed, examples from the NAEP exam never start with a request for the student to *identify*, *interpret*, *represent*, or *recognize*. These words are broadly misused in the standards. On the other hand, mathematical words such as *compute* and *algorithm* never appear and *calculate* occurs only twice.

UNCLEAR LANGUAGE

A majority of the standards include awkward, verbose or imprecise, or meaningless phrases. For example, “Use ratios to describe problem situations” instead of “Use ratios to solve problems.” A grade four standard reads, “Explore properties of paths between points.” What this means for the fourth grade is not made clear. A twelfth-grade standard says, “Evaluate algebraic expressions including polynomials and rational expressions.” What does “evaluate x^2 ” mean? One can evaluate the function at a given value for x , but “evaluate a polynomial” makes no sense mathematically. While the problems with unclear language in individual standards can be minor, the general effect is cumulative and quite negative.

Clarity and Specificity Conclusion

Although the standards offer some guidance about the content to be covered by the assessments, the lack of clear priorities is a serious problem for the standards. Students and teachers will not be able to determine which content is important. What’s more, the language used in a majority of the standards makes reading them difficult. The fact that the standards do not provide clear guidance to users results in a score of one for clarity and specificity.

Summary and Grade

The NAEP mathematics standards contain most of the important content we would expect students to learn, but it is hidden within the excessive number of standards (nearly 300). Less important material, such as data analysis, statistics, and probability, receives greater emphasis than more important material like symbolic algebra. Throughout the document, important content standards are bestowed the same status as secondary standards. Often standards containing crucial content are written in a way that also includes minor content, which makes it hard to identify core content.

Over half of the standards would benefit from either major or minor rewriting. Many standards include unnecessary verbiage, mathematically meaningless language, or just a lack of focus. The other major problem—the lack of clear guidance about which content is more important and which is less—is not easy to correct, especially if such a large number of standards are to be retained.

CONTENT	CLARITY	TOTAL	GRADE
5	1	6	C

Trends in International Mathematics and Science Study (TIMSS)

Mathematics

DOCUMENTS REVIEWED

Mullis, Ina V.S., Michael O. Martin, Graham J. Ruddock, Christine Y. O'Sullivan, Alka Arora and Ebru Erberber. 2005. *TIMSS 2007 assessment frameworks*. International Association for the Evaluation of Educational Achievement. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College. http://timss.bc.edu/TIMSS2007/PDF/T07_AF.pdf

Introduction

The Trends in International Mathematics and Science Study (TIMSS) conducts international comparisons of student achievement in mathematics and science. A project of the International Association for the Evaluation of Educational Achievement, TIMSS has administered math and science tests in roughly forty countries every four years since 1995.

The assessment framework reviewed here is a 170-page document covering both mathematics and science, and provides some background information about the tests. Our focus is the nearly thirty pages that cover the TIMSS mathematics standards for the fourth- and eighth-grade tests, and the sample math test items in the framework's Appendix B.

The mathematics framework is organized around content domains (the subject matter to be assessed—i.e., standards) and cognitive domains (“the behaviors expected of students as they engage with the mathematics content”). The content domains are listed separately for fourth and eighth grades, while the cognitive domains are discussed for both grades simultaneously.

The fourth grade has three content domains: number (50 percent), geometric shapes and measures (35 percent), and data display (15 percent). The cognitive domains include knowing (40 percent), applying (40 percent), and reasoning (20 percent). (The numbers listed here represent the percentage of test items falling under each domain.)

The eighth grade has four content domains: number (30 percent), algebra (30 percent), geometry (20 percent), and data and chance (20 percent). The cognitive domains include knowing (35 percent), applying (40 percent), and reasoning (25 percent).

Review

The TIMSS math framework is evaluated on two dimensions, “content and rigor” and “clarity and specificity,” and this review addresses these dimensions in several sections below. First, the content covered by the TIMSS framework is compared with the content that should be covered by a mathematics test for fourth- and eighth-graders (see “Math Content-Specific Criteria,” on page 11). A separate section describes the content covered by the TIMSS framework for each grade level, as well as what is missing from and problematic about it. The next section sums up the content and rigor of the TIMSS framework across both grade levels. Then the review appraises the clarity and specificity of the framework as a whole, and the last section offers an overall summary and final grade.

In sum, the TIMSS standards are a model of clarity. Though they have a few gaps in content, the standards generally cover what is there well and provide sound guidance to readers.

Fourth Grade

CONTENT COVERED

Whole-number arithmetic and problem solving should make up the bulk of learning in grades one through four mathematics; both are addressed. The measurement of lengths and the computation of perimeters and areas of rectangles are also covered.

Additionally, simple fraction and decimal comparison, including adding simple fractions, are covered. Estimation and the basic vocabulary for geometry and elementary data analysis and statistics are included as well.

CONTENT MISSING

Some important topics are missing, including the single-digit addition and multiplication facts and the corresponding subtraction and division facts; the inverse nature of addition and subtraction and of multiplication and division; the precision of measurement; the basic properties of operations such as commutativity and associativity; and elementary probability. In addition, though the sample questions use some of the addition, multiplication, subtraction, and division facts, the facts included are rudimentary.

CONTENT PROBLEMS

The standards on patterns are easily misinterpreted. One reads, “Extend patterns and find missing terms in them.” Math problems addressing this overly nebulous standard must be carefully worded to avoid having multiple possible ways to extend a pattern.

Eighth Grade

CONTENT COVERED

Grades five through eight should thoroughly cover the arithmetic of rational numbers, including decimals, rates, ratios, proportions, and percentages. With the exception of rates, this material is well covered. In addition, the content standards cover similar triangles, perimeters, volumes, circles, and applications of the Pythagorean Theorem (in the geometry section). Problem solving is included throughout.

Basic data analysis, statistics and probability, and basic linear equations are also well covered.

CONTENT MISSING

Both rates (mentioned above) and absolute value are missing.

CONTENT PROBLEMS

The eighth-grade standards for patterns—like the standards for patterns in fourth grade—are also open to misinterpretation. Again, problems associated with these standards must take care to avoid having multiple possible solutions.

SAMPLE PROBLEMS

The sample problems provided in an appendix to the framework are generally quite reasonable.

Content and Rigor Conclusion

Though the content in both grades four and eight is covered nicely, there is just enough missing—and the pattern standards are problematic enough—to give the framework a content and rigor score of six. (Recall that a six means that no more than 20 percent of important content may be missing.)

Clarity and Specificity

Overall, organizing the framework into multiple content domains for both the fourth- and eighth-grade standards works well, even though the domains are not the same for both grades. The number of standards is also appropriate—thirty-eight for fourth grade and fifty-five for eighth grade. Further, the content domains are organized into helpful sub-domains; for example, “Number” in the grade four standards is divided into four sections: whole numbers, fractions and decimals, number sentences, and patterns and relationships.

This clarity of organization is complemented by clarity of language. In other words, almost all the standards are clearly written and easy to understand. Consider some examples from fourth- and eighth-grade math:

- Compare and order whole numbers.
- Solve problems involving proportions.
- Compute with fractions and decimals.
- Solve problems involving percents and proportions.

Though there are a few awkwardly phrased standards, they are the exception.

The standards for both grades four and eight are models of clarity and appropriate balance. They focus on important material and ignore unimportant material. They provide excellent guidance to students and teachers. They easily merit a score of three.

Summary and Grade

The TIMSS standards for both fourth and eighth grade are clear, coherent, and well organized. They provide solid guidance to all readers and users, from students to textbook writers; they are also measurable and burdened by very little jargon. They cover nearly all of the required content, certainly over 80 percent, at appropriate levels of rigor. Excellent decisions have been made about what content to include and exclude, and there is little ambiguity about what is expected. Minor corrections to the pattern standards and the inclusion of some missing standards would raise the score for TIMSS to a perfect ten. Though the TIMSS architects do not have to address the more complicated world of twelfth-grade standards, standards for fourth and eighth grade are still difficult to do well—and here they have been handled quite admirably.

CONTENT	CLARITY	TOTAL	GRADE
6	3	9	A

Programme for International Student Assessment (PISA) Mathematics

DOCUMENTS REVIEWED

Organisation for Economic Co-operation and Development (OECD). 2006. *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: Author. <http://www.pisa.oecd.org/dataoecd/63/35/37464175.pdf>

----- 2006. *PISA released items – mathematics*. Paris: Author. <http://www.oecd.org/dataoecd/14/10/38709418.pdf>

Introduction

The Programme for International Student Assessment (PISA) is governed and administered by the Organisation for Economic Co-operation and Development (OECD), an entity comprised of thirty countries. The PISA exams cover scientific, reading, and mathematical literacy, and have been administered every three years since 2000. This review covers the framework for the math assessment administered in 2006,⁷ in which fifty-six OECD and non-OECD countries participated.

The PISA test is administered to fifteen year-olds since this is the age “in most OECD countries [where] students are approaching the end of compulsory schooling.” PISA’s objective is to test “not so much in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life.” The PISA framework “defines the contents [sic] that students need to acquire, the processes that need to be performed and the contexts in which knowledge and skills are applied.” Though not intended exclusively for school-based education, this framework thus has the same intention and performs essentially the same function as a set of academic standards, and can reasonably be evaluated by the same standards-based criteria. Further, a number of American educators, policymakers, and standard setters look to PISA as a benchmark for what should be required and/or expected of American schools. Since the PISA standards are used for these purposes, it is important to appraise their content.

The framework devotes over forty pages to summarizing what mathematical literacy should mean for fifteen year-olds. PISA breaks content into four categories: “space and shape; change and relationships; quantity; and uncertainty.” (In addition to those four categories, PISA discusses four “situation types,” eight “competencies,” and three “competency clusters,” but these are vague and do nothing to clarify the content that students should have mastered.) For each of the four content categories, there is a general discussion about the mathematics in the category, but again, these discussions do not describe content to be covered. Each category features a short list of “key aspects” that are as close to standards as the framework gets, but these, too, are vague and non-specific. They total only twenty-three and do not supply clear guidance for readers or users (e.g., teachers, curriculum developers, test developers, mathematicians). Sample problems are also supplied in each category, and the real content to be covered must often be inferred from these examples (as well as from released items from actual PISA exams, discussed below). The sample problems illustrate the very low level of mathematics content that is required. The exam is primarily a problem solving exam, and seldom requires the highest level of mathematics that a fifteen year-old would be expected to know.

The second document reviewed here, the released items, is 106 pages long and contains fifty problems with descriptive names (such as Walking, The Best Car, and Postal Charges). Many of these problems consist of several distinct test items, all related to the same theme. All PISA problems are “in context” (i.e., they are intended as real world problems).

Review

The PISA framework is evaluated on two dimensions, “content and rigor” and “clarity and specificity,” and this review will address these dimensions under four sections. First, the content addressed by the PISA materials is compared with the content

⁷ The 2009 PISA Mathematics Framework is scheduled to be released by the end of 2009. We plan to review it in a subsequent report.

that should be covered by a mathematics test for fifteen year-olds (see “Math Content-Specific Criteria,” page 11), and the results are presented under “content covered” and “content missing.” Specific problems with the content are discussed in the third section, followed by a discussion of the released items in the fourth section. The remainder of the review sums up the content and rigor of the PISA framework and released items and considers the clarity and specificity of the former. The last section provides an overall summary and final grade.

CONTENT COVERED

Any mathematics assessment for fifteen year-olds should thoroughly cover the arithmetic of rational numbers, including decimals, and should also cover rates, ratios, proportions, and percentages. Fractions or rational numbers are not mentioned anywhere, but we do find that understanding the meaning of operations “includes the ability to perform operations involving comparisons, ratios and percentages.” The standard covering “number sense” does explicitly include proportional reasoning.

Under the content category “quantity,” we find some standards that are potentially relevant: “understanding the meaning of operations” and “elegant computations.” These are much too vague, however, to assure coverage of the arithmetic of rational numbers. Typically, standards are self-explanatory. However, PISA standards require one to read the lengthy prose surrounding them to gain some insight into their meaning (more on this below).

Volume, area, and perimeter are mentioned under the category “quantity,” and triangle appears in the preliminary discussion and in an example problem. “Similar” occurs in the discussion, but it is not the technical *similarity* of geometry. Coordinates are also mentioned in the discussion. The most guidance we get in the “space and shape” content category is from the rather vague standard: “recognizing shapes and patterns.” This, in principle, could cover much of the material of geometry, but the standard is far too general to be useful.

Students should have covered a full year of algebra by age fifteen. Under the content category “change and relationships,” we have two potentially relevant standards: “representing changes in a comprehensible form” and “understanding the fundamental types of change.” These do not give clear guidance to readers, though. In the accompanying discussion, we find absolute value, linear equations, and inequalities and linear functions.

Note that the discussion material included in the PISA framework is not an obvious attempt to clarify individual standards. Rather, it is a very *general* discussion about the uses of mathematics in the real world. Here is an example:

Every natural phenomenon is a manifestation of change and the world around us displays a multitude of temporary and permanent relationships among phenomena. Examples are organisms changing as they grow, the cycle of seasons, the ebb and flow of tides, cycles of unemployment, weather changes and stock exchange indices. Some of these change processes involve and can be described or modeled by straightforward mathematical functions: linear, exponential, periodic or logistic, either discrete or continuous. But many relationships fall into different categories and data analysis is often essential to determine the kind of relationship that is present. Mathematical relationships often take the shape of equations or inequalities, but relations of a more general nature (e.g. equivalence, divisibility, inclusion, to mention but a few) may appear as well.

Though PISA is given credit here for covering linear equations and inequalities, this credit might be considered quite generous. For one does not find explicit guidance about content so much as suggestions that certain content should be covered—linear equations for example.

CONTENT MISSING

The core content of the arithmetic of rational numbers is missing, as is any coverage of rates. Most of the expected geometry is missing. There are no similar and congruent triangles; circles; angles associated with triangles and parallel lines; or computation of areas and perimeters of rectangles. The Pythagorean Theorem is missing.

Roots, reciprocals, and powers are absent, as is the arithmetic of polynomials and rational expressions. Factoring is also missing. Quadratics are not mentioned.

CONTENT PROBLEMS

PISA's content category “uncertainty” covers “collecting data, data analysis and display/visualisation, probability and inference.” It states:

Relatively recent recommendations concerning school curricula are unanimous in suggesting that statistics and probability should occupy a much more prominent place than has been the case in the past.

Although publications from 1982 to 2000 are cited to support this sentiment, “unanimous” is a dramatic overstatement. In the sidebar on data analysis, statistics, and probability, “How much DASP do students need?” on page 12 of this report, the case is made that these content areas should receive only limited attention. Simply put, there is an overemphasis on this content category in the PISA standards.

RELEASED ITEMS

In addition to the sample problems included in the PISA Framework, PISA also makes available a set of released items (test problems used in previous years). Because of the imprecise nature of the standards and the discussion surrounding them, it is important to review the released items to see if they give further guidance about content.

The most striking feature of the released items is that thirty-five of the fifty involve the use of a picture, table, or graph. “Data display” falls under the content category of “uncertainty,” and this means that the “uncertainty” content category is highly overrepresented, even if the problems go on to test content covered by other categories.

There are only a few real formulas and equations, and only about 10 percent of the released items use any algebra. There is only one question that expects students to produce a formula of their own from the information given in a problem. This question requires students simply to multiply a given formula by 0.8 in order to arrive at the new formula. The lack of algebra illustrates the low level of mathematical content knowledge expected by the PISA assessment. Nor are arithmetic skills in much demand in the problems. Even where such skills might be useful, calculators are allowed.

The level of geometry used in a few of the released items is significantly higher than that suggested by the standards and the discussion surrounding them (as is the case with similarity and congruence). However, as with algebra, only a few test items call upon this level of geometry.

Most of the released items are focused on problem solving and use fairly low-level mathematics content. However, the problems can be quite complex; about 20 percent of them are multi-step problems. So, even though the content is undemanding, the sample problems can be quite difficult due to the number of steps required to solve them.

A significant number of problems could be taken to task for errors, misleading statements, or imprecise questions. Enumerating all of these issues is beyond the scope of this review.

Content and Rigor Conclusion

Arithmetic, most of the geometry, and much of the algebra that is listed in the “Math Content-Specific Criteria,” (page 11) are missing from the PISA framework. More than half of the important content is never mentioned, which would normally result in a score of three for content and rigor—that is, if the framework were all that’s being evaluated. However, the inclusion of the released items affects PISA’s grade for content and rigor. Even though there are very few geometry problems in the released items, those that do appear require more geometry content than the standards suggest. This raises PISA’s content and rigor score from three to four as it now appears that at least 50 percent of the content is covered, but certainly more than 35 percent is missing (see the “Common Grading Metric,” page 16).

Clarity and Specificity

The actual standards included in the PISA framework are non-specific and not testable. They give almost no guidance to readers and users—i.e., teachers, students, parents, curriculum designers, test makers, textbook developers, standards writers, policymakers, or others. The discussions around the standards are rambling accounts of mathematics in the real world, and although they mention various bits of content (for which coverage has been generously credited in this review), these discussions, too, give scant guidance to readers at any level.

According to the “Common Grading Metric” (page 16), a score of zero for clarity and specificity means:

- | The standards are incoherent and/or disorganized. The standards are not helpful to users. The standards are completely lacking in detail.

It is as if the PISA framework were written to illustrate the kind of standards that should merit a score of zero. Accordingly, its score for clarity and specificity score is zero.

Summary and Grade

The PISA assessment tests mathematical literacy, not knowledge and understanding of grade-level mathematical content. The standards and their explanations do not cover the appropriate grade-level material and the released items indicate that the exam is quite weak in mathematical content. It is a problem-solving test, and although mathematics is used, it is somewhat incidental. Many problems have no apparent mathematical content at all and are, at best, small logic puzzles. Because of this low level of required content knowledge, the claim that PISA tests “preparedness for further study” is in doubt.

The test itself is unbalanced, with glaring overemphasis on data display. Most of the content that is expected of a *fifteen year-old* in PISA is what younger students should have already mastered.

As a serious problem solving test using elementary mathematics, the PISA assessment might function nicely, although its unbalanced nature would limit its usefulness here. Certainly schools should teach problem solving in their mathematics curricula, and it should be a major embarrassment for a country to perform poorly on this test. Still, results from PISA ought not to be used to interpret how successful a school system is at imparting grade-level mathematical knowledge and understanding, nor are the PISA framework and released items a suitable model for U.S. standards setters at any level.

CONTENT	CLARITY	TOTAL	GRADE
4	0	4	D

Reviews: English Language Arts

■ SHEILA BYRD CARMICHAEL

- 1 Common Core State Standards for Reading, Writing, Speaking and Listening
- 2 National Assessment of Educational Progress (NAEP) Reading (2009) and Writing (2011)
- 3 Programme for International Student Assessment (PISA) Reading Literacy

Common Core State Standards for Reading, Writing, Speaking and Listening

DOCUMENTS REVIEWED

Common Core State Standards Initiative. 2009. *College and career readiness standards for reading, writing, and speaking and listening: Draft for review and comment*. Washington, D.C.: National Governors Association Center For Best Practices and Council of Chief State School Officers. September 21. <http://www.corestandards.org/Standards/index.htm>

Introduction

Developed by representatives from the ACT, the College Board, and Achieve, Inc. for the National Governors Association and the Council of Chief State School Officers, the Common Core standards identify “essential college- and career-ready skills and knowledge in reading, writing, and speaking and listening across the disciplines” for students at the end of high school.⁸ The authors suggest that “students who meet these standards will have the...skills to flourish in the diverse, rapidly changing environments of colleges and careers.”

In many ways, these standards successfully define a set of college- and career-ready skills in the relevant domains. They outline expectations in strands typically represented in state English standards: reading, writing, and speaking and listening. They include a section called “Applications of the Core,” which defines standards for research and media; this is also common in state English standards. Despite these familiar strands, however, the authors acknowledge “that the responsibility for teaching such skills must also extend to the other content areas,” suggesting that these college- and career-ready skills are really designed to be cross-disciplinary.

The draft standards do not profess to encompass the full content of English standards, including literature, as most state standards do. Hence we have evaluated them only in the skill domains that they profess to cover (reading, writing, speaking and listening, and the applications of research and media). The document does, however, include standards (in the reading strand) for reading literature and analyzing literary text, so we *have* examined that treatment.

In reading (and to a lesser extent in other strands), the standards tend to shortchange the importance of content knowledge; instead, they function primarily as utilitarian, instrumental, broadly worded, skills-centered goals that are subject to multiple interpretations. The introduction assures readers that the standards “should be *accompanied* by a comprehensive, content-rich curriculum” but appears not to recognize that a state’s English language arts (ELA) standards serve as the foundation of such a curriculum. Though the drafters intend that states and other users will indeed supplement these standards with ample, high quality content (and in several places signal that intention), this important responsibility is explicitly left for others to shoulder. Indeed the introduction states that “many important decisions about curriculum will necessarily be left to states, districts, schools, teachers, professional organizations, and parents.” Good curricula *could* be developed from these standards, yet poor curricula could just as easily be developed because the standards leave so much room for interpretation.

One very helpful addition are the text “exemplars” included “to illustrate the level of complexity that college- and career-ready students should be able to handle independently by the end of high school.” They are, for the most part, well-chosen and indeed illustrative of a good ELA curriculum. In spite of this useful addition, however, no required or even suggested list of titles or authors is provided. The standards alone are not specific enough to provide sufficient guidance to curriculum developers and

⁸ Common Core State Standards Initiative. 2009. *College and career readiness standards for reading, writing, and speaking and listening: Draft for review and comment*. Washington, D.C.: National Governors Association Center For Best Practices and Council of Chief State School Officers. September 21. <http://www.corestandards.org/Standards/index.htm>, i.

teachers. One is left to hope that others will indeed fill that troublesome void with the quality content that the standards-drafters envision but do not itemize.

Review

“Reading Informational and Literary Texts,” “Writing,” and “Speaking and Listening” are the three strands identified in the Core Standards. Each of these strands is divided into two sections: 1) “standards for range and content” and 2) “standards for student performance.” The document also includes standards for two additional strands, “Research” and “Media,” combined in a section called “Applications of the Core.”

Each strand is reviewed in the “content and rigor” section that follows. Using the “English Language Arts Content-Specific Criteria” for the upper grades (see page 15) and the “Common Grading Metric” (see page 16), a sub-grade for content and rigor is included at the end of that section. Following the “content and rigor” section is a similar discussion of the document’s clarity and specificity (with a sub-grade also awarded). The review ends with a summary and final grade for the document.

Content and Rigor

Overall, the content and rigor of most of the core standards is reasonable (within the limits described above), but in several places, crucial content is missing or is not covered satisfactorily, especially in reading. It is sometimes difficult to distinguish between more and less important content within and among strands, and some content is not as rigorous as it should be for college and workplace readiness. The bulk of this discussion about content and rigor focuses on the reading standards because the other strands (writing, speaking and listening) are stronger.

CORE STANDARDS FOR READING INFORMATIONAL AND LITERARY TEXTS

As in the other two strands, this section describes both “Standards for the Range and Content of Student Reading” and “Standards for Student Performance.” Ostensibly, the former section addresses content and the latter section addresses skills, though only the performance standards outline specific student expectations. Separating content from skills weakens the overall rigor in every strand, but in reading neither the content nor the skills section describes expectations in sufficient detail to provide unequivocal guidance for curriculum developers.

STANDARDS FOR RANGE AND CONTENT (READING)

An introductory “How to Read the Document” section notes that the “range and content” standards in each strand “describe the *contexts* in which college- and career-ready students must be able to read, write, speak and listen,” but in the case of reading, these “standards” take the form of general statements about the qualities and components of text types and reading activities. The “How To...” section further states that “these standards are themselves required and carry equal force,” yet it is impossible to discern how these guidelines could be *required* unless the content outlined here were skillfully intertwined with targeted reading skills. (The reading skills are currently defined only very broadly in the “Standards for Student Performance” section of the strand.)

The authors have made a laudable attempt to describe the kinds of texts they want students to read by explaining the desirable components of five characteristics of texts:

- A Complexity
- B Quality
- C Vocabulary
- D Range (i.e., text type: literature, informational, and multimedia)
- E Quantity

For the most part, though, the descriptions read more like rationales for caring about the five characteristics in the first place. The description of *complexity*, for example, says that students should be able “to comprehend complex texts independently” and that students must “read texts characterized by demanding vocabulary,” among other admonitions that are subject to interpretation. The *quality* explanation says that texts should be “rich in content” and come from “a variety of disciplines.” All of this is praiseworthy as far as it goes, but of course the directive is subjective. U.S. educators also need to note that these standards make no mention of exceptional American literature, American literary heritage, or the importance of understanding

that heritage as a matter of cultural literacy (see Reading Criterion 3, page 15). Other high performing countries are less abashed about giving special deference to the study of their own literature.⁹

Vocabulary notes the importance of things like “encountering and mastering a rich vocabulary.” Some important aspects of vocabulary *content*, such as knowledge of roots and affixes or etymology, are not discussed (see Reading Criterion 1, page 15). Where *quantity* is concerned, students “must have the capacity to handle independently the quantity of reading material...” This description is again subject to interpretation, and no specific benchmarks for the desirable amount or lengths of reading materials is included (see Reading Criterion 4, page 15).

Least satisfactory are the descriptions offered for *range*, where text types are addressed. Instead of defining text types and linking them to the applicable skills, the standards sketch very general rationales for why students should be exposed to various types of text. For example, regarding literature: “By immersing themselves in literature, students enlarge their experiences and deepen their understanding of their own and other cultures.” This certainly may serve as one justification for reading widely, but it is not a standard. It would be more helpful if this section addressed specific works and authors that would illustrate the desired quality and complexity, as has been attempted in the “Illustrative Texts” section. Better still would be to combine the content and skills standards into more specific expectations (yet still allow for teacher choice), as in the following twelfth-grade standards from Indiana:

LIT.3.6

Evaluate the way in which authors have used archetypes (original models or patterns, such as best friend, champion, crusader, free spirit, nurturer, outcast, tyrant, and others) drawn from myth and tradition in literature, film, political speeches, and religious writings.

Example: Evaluate the archetypes or characterizations developed by works such as “A Raisin in the Sun” by Lorraine Hansberry, “The Crucible” or “Death of a Salesman” by Arthur Miller, or Exodus by Leon Uris. Explain how the archetype of the fallen creature or outcast in the banishment of Adam and Eve from the Garden of Eden, may be used to interpret and evaluate the characterizations in Shakespeare’s “The Tragedy of Macbeth” or “Othello” and in the Greek tragedy, “Oedipus Rex” by Sophocles.

Similar rationales for reading informational text and multimedia sources follow, and similar efforts to combine content with skills would render useful revisions here as well.

The document appends exemplars of the kinds of texts that students should read. These mostly well-chosen excerpts reflect high expectations, but it is not clear how these exemplars will carry any real weight. They appear more as advisory recommendations that states, districts, and teachers could choose to ignore.

STANDARDS FOR STUDENT PERFORMANCE (READING)

Eighteen statements comprise the performance standards for reading. They set forth such general reading skills as “discern the most important ideas...” and “determine what is meant by words and phrases in context...” The statements are intended to apply to both literary and informational texts, but that strategy works only at the most basic level. For example, while standards for literary analysis are hinted, as in number six, there is no discussion of specific literary elements and devices (see Reading Criterion 2, page 15).¹⁰

Further, this standard “applies to studying characters in fiction and figures in historical texts,” two very different propositions requiring distinct analytical lenses. Where informational text is concerned, the standards address arguments only obliquely, as in number fourteen.¹¹ Specific structures of arguments and types of reasoning, especially *fallacious* reasoning, are not discussed (again, see Reading Criterion 2). Students are left with no specific benchmarks or vocabulary for understanding and explaining why arguments may or may not be valid, nor how dangerous specious arguments can be.

This conflation of skills for reading literary and informational text will make it difficult for teachers to nurture the unique skills that are necessary for navigating certain text types. It also makes prioritization of content difficult to discern. As a result, the

9 Common Core. 2009. *Why we're behind: What top nations teach their students but we don't*. Washington, D.C.: Common Core, 53. http://www.commoncore.org/_docs/CCreport_whybehind.pdf

10 Standard number six reads: “Analyze the traits, motivations, and thoughts of individuals in fiction and nonfiction based on how they are described, what they say and do, and how they interact.”

11 Standard number fourteen reads: “Evaluate the reasoning and rhetoric that support an argument or explanation, including assessing whether the evidence provided is relevant and sufficient.”

standards are less rigorous than they should be at the end of high school.

CORE STANDARDS FOR WRITING

The writing standards exhibit some of the same difficulties as the reading standards by conflating writing skills across genre types, but they do a better job than the reading standards of acknowledging (at least for informational and argumentative writing) the ways in which content and skills must be combined to define rigor.

STANDARDS FOR RANGE AND CONTENT (WRITING)

As in reading, these “standards” describe a set of characteristics of and contexts for writing. They emphasize the “when and why” of student writing rather than delineate specific student expectations for the content expected in various writing applications (genres).

The characteristics are:

- A Purpose (to make an argument or to inform/explain)
- B Audience
- C Situation (on demand writing or writing over time)
- D Technology and Collaboration
- E Quantity

The *purposes* are limited to making an argument and informing/explaining. Some may wonder why narrative writing is marginalized in a side note and why creative writing is not mentioned at all, but it appears that the authors chose to focus on the types of writing most common in both college and workplace settings. One must hope that narrative and creative writing will be addressed at earlier grades in the back-mapped standards to follow.

The other range and content standards for writing provide explanations about when and why students must write often and well, but do not outline specific student expectations. It is particularly disappointing that they do not offer more detailed content for English language conventions. A footnote discusses their importance and even lists some particular concepts for “formal writing” that could be incorporated into these standards. Again, one hopes that specific standards for the correct use of grammar, usage, and mechanics will be detailed in the back-mapped standards to follow.

STANDARDS FOR STUDENT PERFORMANCE (WRITING)

For the most part, these standards do address components of good writing but, as in reading, they are fairly generic. Without explanation, the document identifies the skills specific only for “writing to inform/explain” and for “writing to make an argument.” It does not specify the skills that are important in other writing genres, which would provide better guidance to teachers. Even the argument description could be improved.

For example, in standard number eighteen, it would be useful to include addressing or refuting counterclaims effectively in addition to “acknowledging” them.¹² Providing samples for writing and detailed scoring rubrics, as for reading, would also help to define and illuminate the expected level of rigor.

CORE STANDARDS FOR SPEAKING AND LISTENING

These spare standards combine speaking and listening and focus on effective communication as an overarching goal. As with writing, the “range and content” standards address the “when and why” of speaking and listening, offering as contexts:

- A Group and One-to-One Situations
- B Varied Disciplinary Content (i.e., students must adapt communication skills for different disciplines)
- C Multimedia Comprehension

These contexts are appropriate but, as with the other strands, the statements are not really standards so much as rationales for the importance of speaking and listening skills in these situations. We may infer that they describe when and why one might use

¹² Standard number 18 reads: “Acknowledge competing arguments or information, defending or qualifying the initial claim as appropriate.”

the skills described under “standards for student performance,” but they do not describe content. It seems odd that it is here (B) where the importance of developing a foundation of disciplinary knowledge and conventions is highlighted. Such a discussion is necessary for reading and writing as well, indeed more so.

The performance or skills standards themselves are straightforward and useful, such as “Follow the progression of the speaker’s message, and evaluate the speaker’s point of view, reasoning, and use of rhetoric.” Missing are essential standards for formal oral presentations and recitation—and explanations for their evaluation in the form of detailed scoring rubrics that would help define and illumine the expected level of rigor (see Listening and Speaking Criterion 9, page 15).

APPLICATIONS OF THE CORE

The “Applications of the Core” section describes “the essential skills and knowledge that students need to apply to college and career tasks such as research and media.” Rationales for inclusion of these applications are reasonable. In the case of research, the standards are actually standards from other strands presented in a coherent permutation of steps for formulating research questions, gathering information, analyzing research sources, and reporting findings. Perhaps in a subsequent draft the authors should try a similar experiment with combining skills (and content standards) in reading and apply them to specific types of texts.

The standards for media include both standards for “range and content” and “standards for student performance.” These standards suffer from similar shortcomings as described above. They are not so much standards as descriptions of the kinds of multimedia that graduates will encounter and the contexts in which they will find them. The standards for student performance are taken from other strands—as for research—and presented as a new permutation for application to multimedia. They are fine for the most part, but would benefit from greater detail about the methods for presenting information, especially for combining written, visual and aural components—and identifying their unique effects, such as bias and propaganda in political or other advertising (see Media Criteria fourteen and fifteen, page 16). Including specific standards for developing multimedia presentations might also be useful, as it is an increasingly essential college- and career-readiness skill.

Content and Rigor Conclusion

These standards reflect laudable attempts to identify certain cross-disciplinary skills (typically taught in English classes) and explain why they are necessary for success in college and the workplace. The skills described are mostly reasonable but, because the standards focus on skills, some crucial content—particularly in the essential area of reading—is left unaddressed or is poorly addressed. The skills statements in reading try too hard to serve two masters, informational and literary text, and end up serving neither effectively. Although strong exemplars of the quality and complexity of expected reading are included, they function as suggestions rather than as requirements. This places a very heavy burden on others—including those who will develop the assessments needed to give these standards traction—to establish the nature, quality and content of what should actually be read (and how well). For these reasons, the standards earn a five for content.

Clarity and Specificity

The core standards are mostly succinct, and their organization is relatively straightforward, but they lack coherence and clarity. The prose is generally comprehensible, but it is vague or convoluted in many places, leaving student and teacher expectations unclear. One primary example is the use of the word “standards” to refer to statements that are not standards, as is consistently the case with the “standards for range and content.” These statements are mostly rationales for why certain reading genres or characteristics of writing are important. Instead of outlining content in the genres, as they could, the developers suggest when and in what contexts graduates might need to call on certain reading, writing, speaking and listening *skills*. Two revisions could address this shortcoming: 1) strengthen the “standards for range and content” by including more specific content knowledge and 2) combine these new-and-improved content standards with the skills described in the student performance sections.

In some places, the language of the performance standards (skills) could be much more specific and therefore clearer. For example, performance standards for reading include these three standards, back-to-back:

- (3): Discern the most important ideas, events, or information, and summarize them accurately and concisely.
- (4): Delineate the main ideas or themes in the text and the details that elaborate and support them.
- (5): Determine when, where, and why events unfold in the text, and explain how they relate to one another.

The subtle differences and overlaps among these standards make it difficult to discern the gist and priority of each. More precise

language would clarify whether the first standard is meant to focus on the skills of accurate and concise summary, and if the second is designed to underscore the skill of understanding how supporting details relate to main ideas (and, if so, how that differs from the expectation in number five). Skills in the other strands exhibit some of the same confusion.

Finally, the purpose of some sections remains unclear. For example, the document includes an introductory section called “Student Practices in Reading, Writing and Speaking and Listening.” These practices are said to “undergird and help unify the rest of the standards document,” but it is unclear how they achieve that goal. The statements under each “practice” do not always align with their stated purpose, as in number two, “They Build Strong Content Knowledge.” The statements here read much like a continuation of the broadly worded description of students demonstrating “independence” as learners in number one. In short, the standards are somewhat lacking in coherence, clarity, and organization. For this reason and others, the standards earn a two for clarity and specificity (see “Common Grading Metric,” page 16).

Summary and Grade

These standards do a praiseworthy job of defining essential competencies in reading, writing, and speaking and listening to succeed in both college and the workplace. The authors are to be commended for not falling prey to spurious postmodern theories that disavow close reading and encourage interpretations of a text based solely on how it makes the reader *feel*. Further, the document properly acknowledges that essential communication skills must be embraced and addressed beyond the English classroom, which could lead to valuable collaboration among teachers and more consistent expectations across subjects.

Although it is not the stated goal of these standards to do more than delineate “college- and career-ready skills and knowledge,” there is a risk that they will become *de facto* *English* standards in the states that adopt them. Requiring states to ensure that *these* standards comprise as much as 85 percent of the state’s *English* standards would, if obeyed, imbue this document with responsibility to do more than just outline utilitarian skills that are directly applicable in some college courses and in the workplace.¹³ Perhaps because they anticipated this reality, the authors *have* included standards for reading and analyzing literary texts. Indeed any proper state standards for English must do that and more. They must provide enough curricular guidance to help teachers instill not just useful skills, but also imagination, wonder, and a deep appreciation for our literary heritage. In this way, a strong foundation in English language and literature is an essential component of a content-rich, liberal arts curriculum, which is necessary for the preservation of a free society and robust culture. State standards should help cultivate not only successful college students and workers, but “citizens,” as Dana Gioia has suggested, “who are awakened not only to their own humanity but also to [the] many possibilities of the human world they are about to enter.”¹⁴

Despite their many good qualities described above, these skills-based competencies do not yet serve as a strong framework for states to build the robust liberal arts curricula that will prepare their young people to thrive as citizens in a free society. States adopting these core standards must, therefore, be very careful about how they supplement them so as to achieve that goal.

CONTENT	CLARITY	TOTAL	GRADE
5	2	7	B

13 Common Core State Standards Initiative. 2009. *Executive Summary*. Washington, D.C.: National Governors Association Center for Best Practices and Council of Chief State School Officers. <http://www.corestandards.org/Files/CCSSIOne-Page.pdf>

14 Gioia, Dana. 2007. “Pleasure, Beauty and Wonder.” *Beyond the Basics: Achieving a liberal education for all children*. Washington, D.C.: Thomas B. Fordham Institute.

National Assessment of Educational Progress (NAEP) Reading (2009) and Writing (2011)

DOCUMENTS REVIEWED

National Assessment Governing Board. 2008. *Reading framework for the 2009 National Assessment of Educational Progress*. Washington D.C.: U.S. Department of Education. <http://www.nagb.org/publications/frameworks/reading09.pdf>

American Institutes for Research. 2009. *Reading assessment item specifications for the National Assessment of Educational Progress*. National Assessment Governing Board. Washington D.C.: U.S. Department of Education. <http://www.nagb.org/publications/frameworks/reading-specs-2009.doc>

National Assessment Governing Board. 2009. *National Assessment of Educational Progress 2009 sample questions booklet: Grade 4 mathematics, reading, science*. National Center for Education Statistics. Washington D.C.: U.S. Department of Education. http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-O-G04-MRS.pdf

National Assessment Governing Board. 2009. *National Assessment of Educational Progress 2009 sample questions booklet: Grade 8 mathematics, reading, science*. National Center for Education Statistics. Washington D.C.: U.S. Department of Education. http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-G08-MRS.pdf

National Assessment Governing Board. 2009. *National Assessment of Educational Progress 2009 sample questions booklet: Grade 12 mathematics, reading, science*. National Center for Education Statistics. Washington D.C.: U.S. Department of Education. http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/09SQ-G12-MRS.pdf

ACT, Inc. 2007. *Writing framework for the 2011 National Assessment of Educational Progress, pre-publication edition*. National Assessment Governing Board. Washington D.C.: U.S. Department of Education. <http://www.nagb.org/publications/frameworks/2011naep-writing-framework.doc>

ACT, Inc. 2007. *Writing specifications for the 2011 National Assessment of Educational Progress, pre-publication edition*. National Assessment Governing Board. Washington D.C.: U.S. Department of Education. <http://www.nagb.org/publications/frameworks/2011-naep-writing-specifications.doc>

Introduction

Both the NAEP 2009 Reading Framework and the 2011 Writing Framework reflect an unfortunate trend among developers of K-12 academic standards, namely the perceived need to assign utilitarian “real world” purposes to all academic expectations. The NAEP 2009 reading framework openly acknowledges the influence of PISA, a proudly *non*-curriculum-based exam, which describes itself as a test that “assesses how far [sic] students near the end of compulsory education have acquired some of the knowledge and skills essential for full participation in society.”¹⁵ Both frameworks were also developed in the wake of the 2004 release of the American Diploma Project (ADP) Benchmarks, which prompted many policymakers to evaluate state standards and high-school graduation requirements strictly in the insufficient light of the knowledge and skills that are *necessary for college*

15 Organisation for Economic Cooperation and Development (OECD). No date. *OECD Programme for International Student Assessment*. Paris: Author. http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html

and work¹⁶—rather than also recognizing their vital role as the foundation of a robust liberal arts curriculum, for responsible adulthood, and for competent citizenship.

Both NAEP frameworks seek to connect the content and skills they delineate to their ostensible application in the postsecondary worlds of college and work. While it is true—and important—that students should be prepared to succeed in college and the workplace, not all valuable educational experiences, such as the analysis of great literature, will have a direct or perpetual application in the workplace. (Some aren’t necessarily applicable in college, either. A student may never take a “poetry course” in college, yet still should have learned to read poetry in high school.) Many things learned in school but not immediately useful after high school nevertheless remain essential for developing a student’s cultural literacy, capacity for empathy, understanding of history, and potential for effective citizenship. Fortunately, these attempted connections to real world applications appear to function largely as rationales and do not completely eviscerate the *substance* of the NAEP assessment frameworks, which remains largely intact compared with previous iterations.

Review

“Reading” and “Writing” are the only strands of English Language Arts assessed by NAEP, and each has its own framework. Thus, the reviews of each framework are presented separately here, with individual grades assigned. The 2009 Reading framework is addressed first, with a discussion of “content and rigor” that is organized into two subcategories: “Coverage of Text Types” and “Types of Reading Skills” assessed. A discussion of “clarity and specificity” follows; the review ends with a summary and final grade. The Writing Framework is organized in similar fashion except there the “content and rigor” section is subdivided into two categories titled “Purposes for Writing” and “Features of Writing.”

Both Frameworks are evaluated against their stated purposes. That is, the NAEP Reading Framework is assessed as a “reading” framework only. It does not address writing, communication skills, research, media, or language, all worthy components of a robust English Language Arts curriculum—but not the territory that NAEP here seeks to occupy. Similarly, the NAEP Writing Framework is evaluated only as a “writing” framework. Therefore, the reading and writing criteria of the “English Language Arts Content-Specific Criteria” (see page 13) are the sole criteria applied to the respective reviews.

Readers should bear in mind that the purpose of NAEP frameworks is to delineate the parameters and content of large scale, matrix sample assessments. In matrix sampling, no one student takes the whole test. Each test-taker responds to different sets of questions, which limits testing time yet still provides a picture of general student performance on a broad range of content and skills.

NAEP Reading Framework (2009)

Content and Rigor

COVERAGE OF TEXT TYPES

The content of the 2009 NAEP Reading Framework focuses on two types of texts:

- **Literary texts** (fiction, literary nonfiction, and poetry) and
- **Informational texts** (exposition, argumentation, and persuasive text; and procedural text and documents).

The framework prescribes the reporting of separate subscales for literary and informational text, which NAEP has never done before. The text types do not, however, differ much from the previous framework, which had defined three text types: “literary, informational, and document.” The document category in the previous framework has been subsumed under the new informational text category, while literary nonfiction (e.g., essays and speeches) has been moved from the informational category to the new literary text category. This redistribution is significant in the context of what the overall passage distribution looks like in the 2009 framework (see Exhibit 1).

¹⁶ Achieve, Inc. 2004. *Ready or not: Creating a high school diploma that counts*. Washington, D.C.: American Diploma Project. http://www.achieve.org/files/ADPreport_7.pdf

EXHIBIT 1. *Percentage distribution of literary and informational passages*

Grade	Literary	Informational
Grade 4	50	50
Grade 8	45	55
Grade 12	30	70

In the twelfth grade, significantly more weight is given to informational text than to literary text.¹⁷ Despite the laudable addition of poetry at fourth grade, the framework maintains a disproportionate emphasis on informational texts. Insofar as NAEP is influential in the states, this shift could result in diminished emphasis in the field on teaching literary texts, both fiction and nonfiction. Further, it could have profound and detrimental effects on students' cultural literacy and their foundational skills in the humanities.

The framework supplies details of the various types of text in easy-to-read matrices. For the most part, they represent a reasonable set of expectations for K-12 students, although there is room for improvement. First, some “aspects of texts” at eighth and twelfth grades might easily be introduced earlier. For example, rhythm patterns, point of view, and personification, now listed at eighth grade, could certainly be included at grade four. Aspects of fiction at grade eight could also be assessed at grade four, such as flashback and personification. Some aspects at twelfth grade also seem to come late, such as monologue, comic relief, and “unconventional use of language.”

Second, the aspects of “argumentation and persuasive text” at twelfth grade are left undefined, despite this subcategory’s previously stated significance. Finally, the elusive notion of voice *is* introduced as early as grade four in sub-categories of informational text, but how or why voice is distinct from style, tone, or mood is a much-debated proposition; the inclusion of voice seems unnecessary.

TYPES OF READING SKILLS ASSESSED

As for what students are asked to do with the texts, the framework defines three types of “cognitive targets, or behaviors and skills” to be assessed:

- ▣ Locate/recall;
- ▣ Integrate/interpret; and
- ▣ Critique/evaluate.

These skill categories do not differ significantly from those detailed in the previous NAEP reading framework, but specific skills are not delineated by grade level, which would help clarify expectations. The framework explains:

The cognitive targets remain the same across all three grades on the assessment but the passages and documents about which items are developed will be of increasing sophistication at each grade.

Because the “targets” do not change across the grade levels, the rigor of the NAEP reading assessment ultimately depends on how particular grade-specific items are constructed and the types of passages that are selected for students' consideration.

Unfortunately, we do not have access to the full range of test items and associated passages at each grade span, but the released items found in “Sample Questions” booklets offer a glimpse of the assessment’s rigor. The sample assessment items at each grade span seem appropriately challenging in terms of the cognitive skills listed above, but the sample reading passages themselves are largely mediocre and not particularly complex. As a result, the sample items overall are not sufficiently demanding.

One welcome addition to the 2009 framework is “a more systematic approach to vocabulary assessment,” including the

17 The emphasis on informational text appears to be due to concerns about “college- and career-readiness.” Indeed the framework’s Executive Summary notes that the National Assessment Governing Board (NAGB) adopted “minor modifications to the 2009 Reading Framework at grade 12 based on a comprehensive analysis of the framework conducted by Achieve, Inc.” The referenced 2005 report (*Recommendations to the National Assessment Governing Board on Aligning 12th Grade NAEP with College and Workplace Expectations: Reading*) states: “Because the reading demands that high school graduates face are overwhelmingly informational in nature and informational literacy is a vital 21st century skill, NAEP should increase the percentage of informational text on the new NAEP from 60 to 70 percent.” NAGB also cites the Progress in International Reading Literacy Survey (PIRLS) and PISA reading definitions as bases for the framework’s rationale; both definitions emphasize what is “valued by” or “necessary” for “participation in society.”

calculation of a vocabulary sub-score for the 2009 administration of the test. Although this new emphasis on vocabulary is commendable, the sample questions send mixed signals about rigor. Fourth and eighth grade look to be reasonably rigorous, but twelfth grade does not. For example, the words “scornfully” and “subdued” at fourth grade and “deflect” and “abated” at eighth grade are of an appropriate difficulty level; “obligations” and “precedent” at twelfth grade are not. Words taken from SAT prep lists, such as “specious,” “attenuate” or “circuitous” might be more appropriate.

Content and Rigor Conclusion

While the framework serves its designated purpose well, namely to be a fair starting point for assessing basic reading comprehension at fourth, eighth, and twelfth grades, it does not do four things that Americans should expect academic standards in reading to do: 1) address expectations for essential early reading skills; 2) detail more explicit aspects of vocabulary development; 3) prioritize students’ exposure to outstanding American literature; and 4) detail choices about the quality and complexity of reading passages.

The framework is missing crucial content by not placing enough emphasis on literature and English language conventions. The framework could also cover the content more carefully by distinguishing the skills targets at each grade span, as well as making them more rigorous and grade appropriate. These revisions would also help practitioners distinguish between more and less important content. Since the framework falls short in these ways, it earns a grade of five (see “Common Grading Metric,” page 16).

Clarity and Specificity

The reading framework scores fairly well against our criteria for clarity. It is generally clear and usable. However, as noted above, it could be more specific and detailed in ways that would also improve the content.

Despite its general clarity, the persistence of some vague language in the framework makes it difficult to give it the highest marks. For example, the definitions of the text types “argumentation” and “persuasion” are confusing. The framework implies that argumentation must, by definition, be persuasive text, and it links the two types together as one sub-category of informational text. Many practitioners, however, would distinguish between presenting an “argument” as a matter of exposition (i.e., making known one’s position on an issue or one’s interpretation of a text) and a persuasive essay or speech (i.e., attempting to sway the reader’s opinion, often by including a call to action). The conflation of the two here is likely to confuse practitioners who have been teaching that not all arguments *must* be persuasive.

The need for more specificity where the content is concerned and the confusion surrounding some definitions result in the framework earning a grade of two.

Grade: NAEP Reading

CONTENT	CLARITY	TOTAL	GRADE
5	2	7	B

NAEP Writing Framework (2011)

Introduction

Like the 2009 Reading Framework, the 2011 NAEP Writing Framework proposes a workforce-oriented rationale for its contents. The framework explains that “the impact of communications technologies has changed the way people write” and asserts that “writing in the 21st century is defined by its frequency and efficiency” rather than by its style, elegance, or even persuasiveness.

More specifically, the framework has the following goals:

- To encourage student writers to move beyond prescriptive or formulaic approaches in their writing.

- To assess grade eight and twelve students' writing using word processing software with commonly available tools.
- To measure students' ability to respond to a writing task in an on-demand scenario.

In short, the framework emphasizes students' ability to make choices about how to respond to a writing task—sometimes using a computer—and do it quickly.

Content and Rigor

PURPOSES FOR WRITING

The 2011 NAEP assessment measures three communicative “purposes” rather than the three “modes” of writing that previous Frameworks sought to gauge. The change is mostly semantic: The ultimate *substance* of the framework—what is assessed—appears relatively unchanged.

The three modes of the 1998-2007 NAEP writing assessment were:

- Persuasive mode:
 - Writing to convince
 - Writing to construct an argument
 - Writing to refute a position
- Informative mode:
 - Description
 - Explanation
 - Analysis
- Narrative mode:
 - First-person and third-person fictional stories, personal essays

The 2011 framework's three purposes are:

- **To persuade**, in order to change the reader's point of view or affect the reader's action;
- **To explain**, in order to expand the reader's understanding; and
- **To convey experience, real or imagined**, in order to communicate individual and imagined experience to others.

The framework offers only very general descriptions of the three purposes. It discusses when and why people write for these purposes, but does not detail specific writing expectations by grade levels as the NAEP reading framework attempts to do.

The relative percentages of items devoted to each purpose seem appropriate (see Exhibit 2).

EXHIBIT 2. *Percentage distribution of communicative purposes by grade*

Grade	To Persuade	To Explain	To Convey Experience
Grade 4	30	35	35
Grade 8	35	35	30
Grade 12	40	40	20

There need not be an emphasis on *writing* literature (i.e., “writing to convey experience”) in the writing framework like there is on *reading* literature in the reading framework. In other words, students should not be expected to write sonnets or great literary pieces. Emphasizing persuasive and explanatory purposes for writing, then, makes sense. The emphasis is also consistent with the content criteria for writing (see “English Language Arts Content-Specific Criteria,” page 13) which state that students should “...produce writing that reflects the defining characteristics of various grade appropriate writing genres.”

Compared to the previous writing framework, the emphasis on persuasive writing has increased slightly at grade four (30 percent in 2011 versus 25 percent in 1998). The framework notes that this addition was deliberate, in order to “scaffold” the emphasis on persuasive and informational writing skills at later grades. The “English Language Arts Content-Specific Criteria”

place similar emphases across the three grade spans, moving from “narration, exposition” as an emphasis in kindergarten through fourth grade, to “argument” in five through eight, and “persuasion” at nine through twelve. The writing purposes and the time accorded to each in the NAEP writing framework are therefore sensible and on-target.

FEATURES OF WRITING

Within these types of writing, the framework identifies three very broad categories of features of writing:

- 1 Development of ideas
- 2 Organization of ideas
- 3 Language facility and conventions

Further criteria for assessing writing in each of these “broad domains” are included, but they, too, are rather vague, mentioning only such aspects as “depth and complexity” (for number one above), “coherence” (for number two) and “voice and tone” (for number three). These terms alone are too broad to be useful. Further clarification is attempted in later prose descriptions, but alas, these, too, are nebulous and lack sample responses. For example, when describing “Approaches to Thinking and Writing,” an aspect of “Development of Ideas,” the framework provides this befuddling statement:

Specific approaches to thinking and writing will not be specified on NAEP tasks, but responses will be evaluated for their use of effective approaches in relation to the development and organization of ideas.

The framework appears to rely on sample writing tasks (or “prompts”) more than on these descriptions to exemplify its expectations. This is to be expected since one of NAEP framework’s primary roles is to inform assessment (see “NAEP Overview” in the Appendix, page 52).¹⁸ Even more helpful, however, would be the addition of sample acceptable responses to the tasks (see related “Clarity and Specificity” discussion below).

The prompts generally provide context and clear directions and many are text-based in some way (meaning the tasks are grounded in specific reference points and therefore provide a “level playing field” for responding). In terms of actual grading, however, the six-point scoring rubrics are “holistic.” This means that specific weights are not assigned to each of the criteria (many of which address the features of writing discussed above), making it difficult to discern where students are doing well and where they are struggling. For example, a student’s response could be well-organized, yet inaccessible because of poor sentence structure and mechanics; another response might boast impressive style, but lack development of ideas. The holistic rubrics do not differentiate among these priorities; they could be made more useful if they were prompt-specific and accompanied by sample acceptable responses.

One overall disappointment is the very general nature of the expectations for English language conventions. (They are absent from the reading framework altogether, even though careful analysis of sentence structure and grammar certainly aids reading comprehension.) Making matters worse, English language conventions—the most consistently noted concern of employers and postsecondary faculty—can easily be overlooked with holistic scoring. NAEP should find a way to articulate its expectations in this critically important area in much greater detail.

Content and Rigor Conclusion

Some of the content in the NAEP Writing Framework is covered satisfactorily. Specifically, it does an adequate job of identifying the appropriate purposes for writing and placing reasonable emphases on each purpose at the various grade levels.

While the majority of the appropriate content is covered by the standards, however, it is shortchanged in important ways. First, more attention to conventions would help make the framework as rigorous as it should be. Second, since the rubrics repeat much of the same broadly-worded expectations from grade to grade, it is difficult to distinguish between more and less important content. It is also possible that the standards need to be more or less rigorous than they are at certain grade levels (i.e., expectations are too high or too low) but, given the one-size-fits-all expectations, it is impossible to tell. Because of these limitations, the writing framework receives an overall content and rigor grade of five (see “Common Grading Metric,” page 16).

¹⁸ In fact, NAEP includes accompanying assessment specifications (i.e., a detailed set of instructions for test developers that are also released to the public). The specifications describe the construction of the prompts themselves, such as requesting that a specific audience be named or clearly implied in the tasks, or that visual stimuli such as graphs, editorial cartoons or illustrations, be used to “encourage student engagement.”

Clarity and Specificity

The writing framework scores rather well against our criteria for clarity. The guidelines for writing tasks are straightforward and coherent, and, as mentioned above, the sample prompts help clarify expectations. That said, illuminating a scope and sequence of writing skills by grade span would provide better guidance to educators. More specificity at these important benchmark grades would also help indicate what expectations should look like in the intervening grades.

The framework's generic scoring rubrics could be strengthened by making them specific to a prompt. The framework could also include sample "anchor papers" (i.e., examples of student responses that are used to train scorers) so that practitioners can understand how the rubrics are applied (e.g., whether one area of the rubric appears to be more heavily weighted in scoring than another).

Finally, rather than simply sketching the purposes for writing and offering generic rubrics for scoring the items, the framework could offer a matrix of more specific writing expectations by grade level, as it does for reading. It might mention the kinds of sentence structures, transitions, and stylistic devices it expects to see in writing samples at each grade span. Without offering sample acceptable responses at each cut score, it is difficult to determine the actual level of rigor of the items. This lack of specificity results in a two for clarity and specificity.

Grade: NAEP Writing

CONTENT	CLARITY	TOTAL	GRADE
5	2	7	B

Summary and Grades: NAEP Reading and Writing

Policymakers could easily draw upon the conceptual frameworks delineated by NAEP for both reading and writing when developing national or state standards for English language arts. However, any such effort must keep in mind that actionable standards (at whatever level they are developed) must ultimately address with specificity *all* the domains of English language arts, and at each grade level, in order to be useful. Such an effort must, in particular, focus more than these frameworks do on literature (especially outstanding U.S. literature) and English language conventions. Using these NAEP frameworks as models for state standards *might* maximize students' performance on NAEP. But it may also lead to an English curriculum that slights both literature and English language conventions, jeopardizing students' chances of developing a strong foundation for a broad liberal arts education.

5	2	7	B	<u>Reading</u>
5	2	7	B	<u>Writing</u>

Programme for International Student Assessment (PISA)

Reading Literacy

DOCUMENTS
REVIEWED

Organisation for Economic Cooperation and Development. 2009. *Reading literacy: A framework for 2009*. Paris: Author.¹⁹

Introduction

The Programme for International Student Assessment (PISA) is a large-scale international survey of reading, mathematics, and science “literacy” administered to fifteen year-olds every three years and sponsored by the Organisation for Economic Cooperation and Development (OECD). Forty-three countries participated in the first PISA administration in 2000 and participation has since grown: in 2009, sixty-two countries took part, including China (for the first time). PISA results have become important to policymakers as a way of comparing student performance across countries in an increasingly competitive international economy, but performance comparisons are only the beginning. The National Governors Association has noted that “many nations are going beyond performance to *benchmark* [emphasis added] their policies and practices with the world’s top performers [on PISA].”²⁰ However, advocates of “international benchmarking” need to be aware that the purpose of PISA is quite different from that of American standards-based assessments.

Standards-based assessments in the U.S. are designed to gauge the extent to which students have mastered specific content and skills delineated in state academic standards. Their purpose is clearly summative in nature, assessing what students have (or haven’t) learned. The purpose of PISA, on the other hand, is “forward looking,” according to the OECD, focusing on students’ ability “to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered a specific school curriculum.” Further, the OECD suggests:

This orientation reflects a change in the goals and objectives of curricula themselves, which increasingly address what students can do with what they learn at school and not merely whether they can reproduce what they have learned.²¹

Whether PISA actually reflects this change and/or has helped to fuel a growing disdain for traditional English curricula—curricula that include the study of literature—is a question for debate, but the *influence* of PISA is no longer debatable. The PISA framework, for example, is one of three primary sources cited by the National Assessment of Educational Progress (NAEP) as “offer[ing] support to the definition for reading advocated in the 2009 NAEP Reading Framework.”²² It is also cited as a source of “evidence” for a number of the Common Core reading/ writing/speaking/listening standards and the mathematics standards currently under development by the National Governors Association and the Council of Chief State School Officers (the first public drafts of which are also reviewed in this report).²³

Each PISA administration emphasizes one of the three literacy areas (math, science, or reading). In 2009, the focus was on

19 The OECD generously agreed to provide Fordham with an advance copy of the PISA 2009 reading framework (as yet unreleased). The framework has been approved by the OECD member governments and, according to OECD, will be available to the public online at www.pisa.oecd.org on December 5th, 2009.

20 National Governors Association, the Council for Chief State School Officers, and Achieve, Inc. 2008. *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, D.C.: Authors. <http://www.nga.org/Files/pdf/0812BENCHMARKING.PDF>, 6

21 Organisation for Economic Cooperation and Development (OECD). 2009. *PISA: The OECD Programme for International Student Assessment*. Paris: Author. <http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf>, 57

22 National Assessment Governing Board. 2008. *Reading framework for the 2009 National Assessment of Educational Progress*. Washington D.C.: U.S. Department of Education. <http://www.nagb.org/publications/frameworks/reading09.pdf>, 6

23 See the Common Core State Standards Initiative website for relevant links to PISA: <http://www.corestandards.org/readingmain.html>.

reading. The new framework outlines a definition of “reading literacy” that includes what PISA calls “reading engagement and metacognition.” It emphasizes “motivational attributes and behavioral characteristics of students’ reading.”²⁴ Because the focus of the PISA framework is on defining how reading skills may be used to meet “real-life” challenges, rather than delineating desired content for students, its usefulness as a model for national and state reading standards is limited—and its growing influence as a source for standards developers in the U.S. warrants serious scrutiny.

The PISA framework describes the conceptual basis for an assessment, as well as the types of texts and tasks that students will encounter—and how the tasks will be scored. So, while PISA is not a curriculum-based or school-based framework, American educators, policymakers, and standard setters often look to it as a benchmark for what they should require and/or expect of American schools. Since PISA is being used for these purposes, it is important to examine its content carefully.

Review

This review appraises PISA as a “reading” framework only, not as a framework for the entirety of English language arts. Missing from PISA are standards for writing, communication skills, research, media, and language—all worthy components of a robust English language arts curriculum, but apparently not within PISA’s scope. Therefore, because PISA is a reading literacy assessment framework, the reading components of the “English Language Arts Content-Specific Criteria” are the only criteria used to evaluate it (see page 13).

The PISA framework is evaluated on two dimensions, “content and rigor” and “clarity and specificity.” The content and rigor discussion is subdivided into four sections that mirror the organization of the PISA framework: 1) defining the conceptual basis for PISA’s definition of reading literacy; 2) the organization of the reading literacy domain; 3) the types of skills required for addressing the tasks; and 4) task excerpts. Next, the clarity and specificity of the framework are appraised as a whole and the concluding section provides an overall summary and final grade.

In short, although assessment frameworks must include good direction for test developers, the PISA framework is confusing in its detail and prioritizes the evaluation of metacognitive reading “strategies” over reading comprehension content.

Content and Rigor

The OECD notes that the PISA frameworks take “an innovative approach to assessing knowledge, skills and attitudes by identifying the key concepts and processes that underpin each subject, rather than merely examining curriculum content.”²⁵ In four opaque sections, the framework:

- 1 defines the conceptual basis for its definition of reading literacy;
- 2 explains the organization of the reading literacy domain (i.e., the kinds of texts and tasks sampled on the assessment);
- 3 describes the types of skills required for addressing the tasks; and, finally,
- 4 offers excerpts from sample tasks.

Each of the components is described and appraised below.

DEFINING THE CONCEPTUAL BASIS FOR PISA’S DEFINITION OF READING LITERACY

The PISA framework states that the goal of education “has shifted from the collection and memorization of information only to encompass a broader concept of knowledge.”²⁶ It adds that “the meaning of knowing has shifted from being able to remember information to being able to find and use it,” which is obviously a disputable contention that drinks deeply from the well of so-called 21st Century Skills.²⁷ In fact, it would be difficult to conjure educators who ever thought that the goal of education was simply to “collect and memorize information” in the first place, so the premise is faulty. Epistemologists would likewise recoil at the suggestion that “to know” is simply to “find and use” information, but these are the PISA assumptions.

24 OECD. 2009. *Reading literacy: A framework for 2009*. Paris: Author, 57

25 ——. No date. “PISA: the OECD Programme for International Student Assessment.” Paris: Author, 6. <http://www.oecd.org/dataoecd/51/27/37474503.pdf>, 6

26 ——. 2009. *Reading literacy: A framework for 2009*. Paris: Author, 7

27 For a discussion of the importance of content knowledge as it relates to reading comprehension, see Willingham, Daniel. 2006. “How knowledge helps: It speeds and strengthens reading comprehension, learning—and thinking.” *American Educator*, 30 No. 1. Spring. http://www.aft.org/pubs-reports/american_educator/issues/spring06/willingham.htm

In addition, the framework asserts that “reading-related skills, attitudes, interests, habits and behaviors have been shown in a number of recent studies to be strongly correlated with reading proficiency.” The evidence cited is the OECD’s own finding that this correlation was stronger than the correlation between proficiency and socioeconomic status in the PISA 2000 administration. In short, the OECD seems more interested in *why* students perform well on PISA than in simply determining *whether* the students possess the requisite reading comprehension skills to answer questions correctly.

THE ORGANIZATION OF THE READING LITERACY DOMAIN

PISA’s organization of the reading literacy domain is arguably the most complicated part of this complex framework. Assessment tasks are categorized by “situation” (private, public, occupational, or educational); “text” (medium, environment, format, and type); and “aspect” (access/retrieve, integrate/interpret, reflect/evaluate, and complex). In laymen’s terms, this means that PISA categorizes its assessment tasks first by the situation in which the task might present itself (i.e., reading for pleasure, for school, etc.); then by text type (i.e., description, narration, exposition, etc.); and finally by skill (i.e., the type of cognitive skill demanded by the item, such as interpreting or evaluating information).²⁸ The framework notes that the categories may overlap.

One important drawback to this organization of the reading literacy domain is its heavy emphasis on non-educational categories. For example, with regard to “situation,” only 28 percent of print text items and a mere 12 percent of electronic text items represent “educational” situations.²⁹ Forty-four percent of print items and 52 percent of electronic items are “personal,” “occupational” and “public” (meaning they are “usually associated with out-of-school settings for children”). While it is true that PISA seeks to assess students’ ability to participate in society, a study that neglects educational tasks would surely thwart such a goal.³⁰

ASPECT: THE TYPES OF SKILLS REQUIRED FOR ADDRESSING THE TASKS

The framework’s discussion of the kinds of skills that each task assesses is relatively straightforward. In fact, the PISA categories are consistent with common assessment classifications (such as those used in NAEP): access and retrieve; integrate and interpret; and reflect and evaluate. However, a fourth category, “complex,” is added for electronic text only; the framework suggests that none of the existing three skill types suffices to describe the skills demanded in electronic text (more on this below).

The tasks include both multiple choice and constructed response items. Partial credit is awarded in some cases for constructed response items. “Access and retrieve” is much like “locate and recall” in NAEP: a simple skill that requires no complex inferences. “Integrate and interpret” parallels the next NAEP category exactly and “focuses the reader on relationships within a text.” Tasks that require students to “reflect and evaluate” ask them “to draw primarily on knowledge outside the text and relate it to what they are reading.”

The framework includes examples of each type of task and makes distinctions about how these skills are employed when electronic text is involved. For example, the designers of the framework seem interested in exploring the various ways in which readers might *access* information (i.e., knowing where to click in an electronic text versus knowing what paragraph to look for in print text). In neither case is the priority simply assessing readers’ comprehension skills.

The distribution of the print tasks by *skill* seems even-handed (50 percent of the items are “integrate and interpret”; 25 percent are access and retrieve; 25 percent are “reflect and evaluate”), yet the distribution by *level of difficulty* is extraordinarily lopsided. None of the print items is deemed “complex”—not even one. This seems to be a task type reserved for electronic media. By restricting skill complexity to electronic media only, the framework implies that *only* electronic texts allow students to access, select, and retrieve information “from a wide field of possible resources.” Yet print media obviously may afford the same opportunities.

TASK EXCERPTS

As with any assessment framework, the sample reading passages, test items, and answers are the best indication of the exam’s rigor. Unfortunately, the 2009 PISA Reading Framework has few sample task items. They include the *description* of an excerpt (i.e., not an actual excerpt) from Gabriel García Márquez’s *One Hundred Years of Solitude* and two reproductions of

28 “Skill” is also addressed in (3) “Aspect.”

29 Recall that assessment tasks are categorized by “situation” first, which is broken down into private, public, occupational, or educational situations.

30 Curiously, the PISA framework also classifies literary text as “personal” since its architects say the “original” purpose of such text is reading for pleasure. If these were the only cases where non-academic text was highlighted, it might not be a problem. As it stands, the “personal” category includes “personal letters” and “texts that are intended to satisfy curiosity, as a part of leisure or recreational activities.” Moreover, electronic text “includes personal emails, instant messaging and diary-style blogs.”

“Metrotransit” maps, accompanied by instructions that ask students to use the maps to draw the shortest distance between two points. These meager examples do not provide much information about the tasks.

In another excerpt, the stimulus for an open-ended response question is provided, but not the “five short argumentative texts” that the student would need to have read in order to answer the question, making it impossible to determine the item’s rigor. Students are asked to state which of the five argumentative texts they agree with and why, citing their own opinion and the main ideas presented. The prompt provides just three lines for a response, and the sample answers are one or two sentences each, such as, “I agree with him because he is concerned with the environment and he thinks we should leave space alone.” The question begs for a longer and more defensible response than that provided by the rather simple-minded sample correct answer. In fact, the sample response lacks substance and fails to demonstrate the ability to make an argument.

Noting the OECD’s need for “more differentiation at the lower and higher ends of student achievement,” the samples deliberately include one “very easy” and one “very difficult” item. The “very easy” item simply asks students to name the company discussed in a product recall notice; the “very difficult” item asks students to discuss the possible purpose of Pericles’s speech from *The Peloponnesian War* about democracy in Athens. We are told that the latter item is “one of the most difficult administered in the 2009 PISA field trial,” yet it is the kind of item that high school English teachers might expect many of their students to navigate successfully. The fact that it is identified as the most difficult suggests that the overall rigor of the PISA assessment is not particularly high. It also suggests that, regardless of what “real world” skills PISA may be assessing, it is not evaluating the menu of knowledge and skills that students should be mastering in English classes before graduating from U.S. high schools.

Content and Rigor Conclusion

As an assessment of “real world” reading literacy, PISA does not meet many of the criteria that characterize rigorous English language arts standards (see “English Language Arts Content-Specific Criteria,” page 13). Even judged purely as a reading framework, which is how it has been assessed here, crucial content is missing: The PISA framework does not address vocabulary development, offer specific expectations for reading and analyzing literary and non-literary texts, include specific expectations for the correct use of English language conventions, recognize the importance of literary heritage, nor sufficiently define the amount, quality, and complexity of literary and informational texts. Further, much of the content of the PISA framework, focused on metacognitive strategies for accessing information, does not address reading comprehension and arguably doesn’t belong there. This means that roughly 50 percent of reading content is missing, resulting in a score of three for content and rigor (see “Common Grading Metric,” page 16).

Clarity and Specificity

The 2009 PISA Reading Literacy Framework is both dense and murky. It includes multiple (and unnecessary) levels of detail about task types. Although these descriptions might be helpful to test item writers, they are not particularly useful as guides for developing national or state standards or for informing curriculum and instruction, even for the age level targeted by PISA (fifteen year-olds). Without any graphical explanation, these multiple layers of description make the framework very difficult to follow. Although the tasks themselves are described with excessive detail, the sample items are incomplete, making it difficult to understand exactly what is expected of students.

Because much of the language in the framework does not provide sufficient guidance to users nor detail about what is expected of students, it receives a score of one (see “Common Grading Metric,” page 16).

Summary and Grade

Policymakers who assume that poor U.S. performance on PISA means that we should make U.S. standards and curricula more PISA-like are making a risky leap. While we should not ignore the results of the PISA reading assessment, we must be cautious about the inferences we make from it, ever mindful that this test and its framework are not—nor do they profess to be—a model for reading standards, much less a complete English language arts curriculum.

CONTENT	CLARITY	TOTAL	GRADE
3	1	4	D

Closing

■ CHESTER E. FINN, JR. AND AMBER M. WINKLER

Like many, we're eager to get high-quality common standards into the hands of state, district, and local officials. We're eager for teachers to start using them and for youngsters to benefit from their use. Truth be told, we're also eager for our kids—and the adults they will become—to give those in Singapore, Hong Kong, and Finland a run for their money.

But we won't allow eagerness to ruin a fifty-years-in-the-making opportunity. It's imperative that America get this done right. That means that states need to do their part to amplify the literary components of the Common Core Standards in reading, writing, listening and speaking. It also means that we must grapple with how to make these common standards stick and get traction—stick as an organizational entity, stick with the post-secondary and workplace worlds, and stick with policymakers, parents, principals, and teachers.

This interim report finds that in late 2009 the United States has got some bright stars by which to navigate (as well as some solar dust to avoid). We need to keep our sights set on them even as gravitational pulls seek to lead us astray.

But, done right, this could become quite a cosmos.

Appendix: Overview of the Standards

■ STAFFORD PALMIERI

This section provides an overview of the four standards/tests that are reviewed in this report: the draft “Common Core” College and Career Readiness Standards; the National Assessment of Educational Progress (NAEP) in English language arts and math; the Trends in International Mathematics and Science Study (TIMSS) in math; and the Programme for International Student Assessment (PISA) in English language arts and math.

“Common Core” College and Career Readiness Standards for Reading, Writing, Speaking and Listening, and for Mathematics

WHAT ARE THEY?

The College and Career Readiness Standards for Reading, Writing, Speaking and Listening, and for Mathematics, currently in draft form, are products of the Common Core State Standards Initiative (CCSSI)—a joint project of the National Governors Association (NGA) and Council for Chief State School Officers (CCSSO), in collaboration with Achieve, ACT, and the College Board. In June 2009, forty-seven states, two territories, and the District of Columbia agreed to work together to develop a set of shared or common (some would say national) academic standards for the U.S. (One additional state joined in September, bringing that total to fifty-one.) This is a privately-funded, state-led effort—and the resulting standards are voluntary.

WHAT IS THE PURPOSE OF THE COMMON CORE STANDARDS?

CCSSI, recognizing the drawbacks of inter-state variation and the consequences of poor-quality standards in some states, is developing common standards that they believe will equip students to be career and college ready. “It is time for us as states to challenge the education system and finally answer the question, ‘What will it take for every child to be successful?’” explains CCSSO Executive Director Gene Wilhoit. “Fewer, clearer, and higher standards will help us get there.”

WHAT DO THE STANDARDS LOOK LIKE?

The draft standards (released on September 21, 2009) address reading, writing, speaking and listening, and math. These subject/skill areas are the foundation of current accountability systems under No Child Left Behind as well as foundational subjects that provide the basis for other subjects.

The reading, writing, and speaking and listening standards are divided into two sub-categories: “range and content” and “student performance.” While the former has three to five broad categories, such as “vocabulary” (reading) and “audience” (writing)—each explaining the areas in which career-and college-ready students should be competent—the latter are more specific. For example, student writers should be able to “use varied sentence structures to engage the readers and achieve cohesion between sentences.” The standards also include a section outlining how these core skills are applied to research and the media. In both sections, there is an emphasis on students’ discerning accurate information and news from the plethora of information. The standards also include a number of sample texts that the developers consider to be high quality plus notes on how students should read and interact with the texts.

The math standards take a somewhat different form. There are many more topics to address, such as number, quality, equations, etc., and each is explained in depth. The math topic standards delineate content (what “students understand”) and skills (what “students can and do”) in two side-by-side columns. For example, under the mathematical concept *number*, students should understand that “quantities can be compared using division, yielding rates and ratios” while also knowing “when and how to use

standard algorithms, and perform them flexibly, accurately and effectively.” This dual focus runs throughout the math standards. Example tasks are presented in an appendix.

HOW WERE THE STANDARDS DEVELOPED?

The NGA, CCSSO, and Achieve, Inc. collaborated on a report entitled *Benchmarking for Success: Ensuring U.S. Students Receive a World-Class Education*, released in December 2008. That document served as the springboard for the Common Core State Standards Initiative. In January 2009, the NGA and CCSSO founded a National Policy Forum to “share ideas, gather input, and inform the common core standards initiative.” Then, in July 2009, the NGA and CCSSO announced the members of the Common Core State Standards Development Work Groups and the Common Core State Standards Feedback Group, one each for English language arts and mathematics.³¹ The former, comprised of content experts from Achieve, ACT, and the College Board, were charged with initial development of the standards; they were aided by independent facilitators, writers, and advisors. The work groups’ discussions were confidential during the process; now that the draft has been released, state and national education groups have the opportunity to comment until October 21, 2009.

After the comment period ends, the standards will be submitted to a twenty-five-person validation committee for further review.³² The extant draft deals with the end of high school but drafters are beginning work to “backward-map” these standards through the earlier grades of school. These K-12 standards will be released in January 2010. Once both sets of standards are finalized, participating states are expected to submit a timeline and process for approving them in 2010. CCSSO and NGA hope to present both sets of standards—the college and career readiness standards and the K-12 grade-specific standards—to states at the same time for approval.

THE FUTURE OF THE STANDARDS

Much of this remains to be determined. Perhaps most important, while U. S. Education Secretary Arne Duncan has said that the federal government will underwrite the development of new assessments to be aligned with the Common Core standards, how and by whom (and when) that will be done is not yet known.

Sources:

Cavanagh, Sean. 2009. “New standards draft offers more details.” *Education Week*, September 25. http://www.edweek.org/ew/articles/2009/09/30/05standards_ep.h29.html?tkn=NUUFPKkm%2BGMVL5uiY5aed6dpIAGEtUuYXTx

Common Core State Standards Initiative. 2009. *College and career readiness standards for reading, writing, and speaking and listening: Draft for review and comment*. September 21. Washington, DC.: National Governors Association Center For Best Practices and Council of Chief State School Officers. <http://www.corestandards.org/Standards/index.htm>

-----. 2009. *College and career readiness standards for mathematics: Draft for review and comment*. September 21. Washington, DC.: National Governors Association Center For Best Practices and Council of Chief State School Officers. <http://www.corestandards.org/Standards/index.htm>

National Governors Association. 2009. “Common Core State Standards Development Work Group and Feedback Group announced.” July 1. News Release. <http://www.nga.org/portal/site/nga/menuitem.6c9a8a9ebc6ae07eee28aca9501010a0/?vgnextoid=60e20e4d3d132210VgnVCM1000005e00100aRCRD&vgnnextchannel=759b8f2005361010VgnVCM1000001a01010aRCRD>

31 Members of the Work Groups and Feedback Group can be found on the CCSSI website: <http://www.nga.org/portal/site/nga/menuitem.6c9a8a9ebc6ae07eee28aca9501010a0/?vgnextoid=60e20e4d3d132210VgnVCM1000005e00100aRCRD&vgnnextchannel=759b8f2005361010VgnVCM1000001a01010aRCRD>

32 Members of the validation committee can be found on the CCSSI website: <http://www.corestandards.org/Files/CoreValidationCommitteeBios.pdf>

National Assessment of Educational Progress (NAEP)

WHAT IS IT?

The National Assessment of Education Progress (NAEP) is a federally funded U.S. testing program, administered to a stratified random sample of students, which is used to monitor academic achievement in core subjects and key grades at the national and state levels. What we reviewed here are the most recent reading/writing and math “frameworks” for what is known as “main NAEP,” which is different from NAEP’s monitoring of long-term achievement trends.

WHAT IS THE PURPOSE OF NAEP?

NAEP was founded in 1964 as a way to benchmark student learning against a common metric. The first assessment occurred in 1969. Today, it still serves much the same purpose and state-level NAEP (dating to 1988) is often used to judge the rigor of states’ own standards and tests and to compare student learning across state borders.

WHO TAKES NAEP?

The NAEP test is administered to a representative sample of fourth-, eighth-, and (usually) twelfth-graders. Some NAEP testing is done each year but different subjects are on different multi-year samples. Under NCLB, reading and math in grades four and eight are tested every two years. Other subjects are assessed less frequently. Some large cities also participate directly in NAEP but otherwise NAEP yields results only at the national, regional and state level—never at the school, classroom, or individual level.

WHAT DOES THE TEST LOOK LIKE?

The tests are composed of multiple choice and constructed response questions. They are accompanied by background questionnaires that address student demographics and educational attainment. Students complete these “non-cognitive” questions as part of the assessment.

HOW IS THE TEST DEVELOPED AND ADMINISTERED?

NAEP is administered by the National Center for Education Statistics (NCES), an arm of the U.S. Department of Education; it outsources much of the test development and administration to contractors. Policy decisions concerning NAEP, however, are made by the independent National Assessment Governing Board (NAGB), whose members are appointed by the secretary of education. NAGB develops the frameworks and test specifications that lay the groundwork for the test itself.

NAGB was created by Congress in 1988. It is a twenty-six-person bipartisan group of governors, state legislators, local and state officials, educators, business representatives, and members of the general public.

NAGB’s framework development process is multi-faceted, including input from state education officials, content experts, the general public, NCES staff, and a policy advisory staff. NAGB also reviews and approves all of the question items; since 1990, it has also set NAEP achievement levels (i.e., definitions of basic, proficient, and advanced) and the assessment schedule. Through its Assessment Development Committee, NAGB closely monitors the development and production of the assessment framework, the item development specifications, and the background questionnaires that accompany the test. Test questions are developed by NAEP contractors and go through a rigorous review process.

THE FUTURE OF NAEP

There has been some discussion of adding twelfth grade to the State NAEP-tested grade levels. This reflects the growing emphasis on ensuring that students graduate from high school career- and college-ready. There is also support for adding science and writing as required State NAEP subject tests. At minimum, those two subjects will be tested on a trial basis, alternating each testing cycle for the next few cycles.

Sources:

National Assessment Governing Board. 2002. “Framework development policy statement.” Washington, DC.: NAGB. May 18. <http://www.nagb.org/policies/PoliciesPDFs/Assessment%20Development/Framework%20Development.pdf>

----- 2002. “Item development and review policy statement.” Washington, DC.: NAGB. May 18. <http://www.nagb.org/policies/PoliciesPDFs/Assessment%20Development/Item%20Development%20and%20Review.pdf>

-----, 2002. "Review of the National Assessment of Educational Progress policy statement." Washington, D.C.: NAGB. August 3. <http://www.nagb.org/policies/PoliciesPDFs/Technical%20Methodology/Review%20of%20the%20National%20Assessment%20of%20Educational%20Progress.pdf>

-----, 2006. "The future of 12th Grade NAEP Report of the Ad Hoc Committee on Planning for NAEP 12th Grade Assessments in 2009." Washington, D.C.: NAGB. August 4. <http://www.nagb.org/policies/PoliciesPDFs/Improving%20National%20Assessment%20of%20Education%20Progress/The%20Future%20of%2012th%20Grade%20NAEP.pdf>

-----, 2007. "Eligibility criteria and procedures for selecting districts for participation in the National Assessment of Educational Progress Trial Urban District Assessment policy statement." Washington, D.C.: NAGB March 3. <http://www.nagb.org/policies/PoliciesPDFs/Program%20Administration/Trial%20Urban%20District%20Assessment%20Policy.pdf>

National Center for Education Statistics. "NAEP overview." Washington, D.C.: U.S. Department of Education. <http://nces.ed.gov/nationsreportcard/about/>

Vinovskis, Maris A. 1998. *Overseeing the Nation's Report Card: The creation and evolution of the National Assessment Governing Board*. National Assessment Governing Board. Washington, D.C.: U.S. Department of Education. <http://www.nagb.org/publications/95222.pdf>

Trends in International Mathematics and Science Study (TIMSS)

WHAT IS IT?

Trends in International Mathematics and Science Study is an assessment of fourth- and eighth-grade math and science knowledge.³³ However, we only reviewed frameworks guiding the math portion in this report. It was created in 1995 by the International Association for the Evaluation of Educational Achievement (IEA), a sixty-eight-member (as of 2009) independent organization of national education research organizations and government education research agencies; the IEA also oversees the Progress in International Reading Literacy Study (PIRLS).³⁴ The two tests are run by the TIMSS & PIRLS International Study Center at Boston College. The TIMSS assessment is given every four years. (The last administration occurred in 2007, with the next to occur in 2011).

WHAT IS THE PURPOSE OF TIMSS?

TIMSS is a curriculum-based test of student achievement. It is concerned with how the curriculum is defined nationally, implemented in schools and classrooms, and absorbed by students. TIMSS calls these three pieces the intended curriculum (“the mathematics and science that society intends for students to learn and how the education system should be organized to facilitate this learning”), the implemented curriculum (“what is actually taught in classrooms, who teaches it, and how it is taught”), and the achieved curriculum (“what it is that students have learned, and what they think about these subjects”). In the end, TIMSS data address a key question: Are students actually learning what the curriculum says they should be?

The test is also designed to capture achievement trends (though only through four testing cycles so far). A number of countries have been part of TIMSS since 1995 and can evaluate their achievement over that period. These trend data also capture achievement by grade cohort, since fourth-graders who take TIMSS in one testing cycle are eighth-graders in the next cycle.

WHO TAKES TIMSS?

Students at the end of four and eight years of schooling (typically fourth and eighth grade, though not necessarily) in participating countries take TIMSS. The number of participating countries has grown from forty-one in 1995 to more than fifty-nine countries and eight other entities (e.g., regional entities that are tested as if they were countries, such as the state of Massachusetts and Dubai, UAE) in 2007. The U.S. has been participating in TIMSS since its first administration. Almost seventy countries are expected to participate in 2011; forty of those countries will also participate in PIRLS, which is administered every five years, and happens to coincide with TIMSS in 2011.

WHAT DOES THE TEST LOOK LIKE?

Math and science are treated equally on the TIMSS assessment. In 2007, for example, there were 179 mathematics items and 174 science items (359 total) in fourth grade and 215 mathematics items and 214 science items (429 total) in eighth grade. Since there are many more questions that can possibly be answered by each individual student, the test uses a matrix-sampling approach whereby each student answers a subset of the total questions. (NAEP and PISA also use this strategy.)

The test includes both multiple choice and constructed response questions (the latter are graded for content, not writing quality). In addition, students, teachers and principals fill out questionnaires that address home and school environments, instructional activities, topics covered, and teachers’ own training. These questions are used to create background indices that accompany the raw achievement data.

HOW IS THE TEST DEVELOPED AND ADMINISTERED?

Prior to 2003, the International Study Center overhauled both the math and science frameworks to better reflect what was actually being taught in participating countries and to make the test data more useful for them. This process included using the expert advice from content experts as well as intensive country surveys.

The draft frameworks are reviewed by the National Research Coordinators (NRCs), a group of individuals chosen by each

³³ TIMSS 1995 included a twelfth grade (or “end of secondary schooling”) assessment. This was discontinued in subsequent administrations.

³⁴ We plan to include a review of PIRLS in the follow-up report to *Stars by Which to Navigate? Scanning National and International Education Standards in 2009*, which is scheduled for publication in spring 2010.

country to represent it and implement the test within its borders in compliance with TIMSS guidelines. NRCs consult with content experts in their respective locales. Next, the frameworks are reviewed by that testing cycle's Science and Mathematics Item Review Committee (SMIRC), comprised of content experts in both fields. Question development is led by the NRCs, with support from the International Study Center and a number of contractors.³⁵

As of 2007, the frameworks are evaluated every testing cycle to keep up with the changing curricula of participating countries (a list that continues to grow) and advancing content expertise in math and science. IEA is, however, keen to maintain continuity in the tested material over time given the importance of trend data.

Sources:

Mullis, Ina V.S., Michael O. Martin, Graham J. Ruddock, Christine Y. O'Sullivan, Alka Arora and Ebru Erberber. 2005. *TIMSS 2007 assessment frameworks*. International Association for the Evaluation of Educational Achievement. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. http://timss.bc.edu/TIMSS2007/PDF/T07_AF.pdf

Mullis, Ina V.S., Michael O. Martin, Graham J. Ruddock, Christine Y. O'Sullivan and Corrina Preuschoff. 2009. *TIMSS 2011 assessment frameworks*. International Association for the Evaluation of Educational Achievement. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. http://timss.bc.edu/timss2011/downloads/TIMSS2011_Frameworks.pdf

Mullis, Ina V.S., Michael O. Martin and Pierre Foy. 2008. *TIMSS 2007 international mathematics report*. International Association for the Evaluation of Educational Achievement. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. http://timss.bc.edu/TIMSS2007/PDF/TIMSS2007_International-MathematicsReport.pdf

Olsen, John F., Michael O. Martin and Ina V.S. Mullis, eds. *TIMSS 2007 technical report*. International Association for the Evaluation of Educational Achievement. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. http://timss.bc.edu/TIMSS2007/PDF/TIMSS2007_TechnicalReport.pdf

35 These include the Educational Testing Service (ETS) and the National Center for Education Statistics (NCES). A variety of other organizations are involved in data collection, psychometric development, and data analysis, including the IEA Data Processing and Research Center in Hamburg, Germany, and Statistics Canada in Ottawa, Canada.

Programme for International Student Assessment (PISA)

WHAT IS IT?

The Programme for International Student Assessment (PISA) is an internationally standardized assessment in reading, math, and science that was jointly developed by participating countries and is administered to fifteen year-olds worldwide; we reviewed the frameworks for reading and math in this report.³⁶ Launched by the Organisation for Economic Cooperation and Development (OECD) in 1997, PISA provides international data on student learning. The first PISA test was administered in 2000, with subsequent tests occurring every three years. Data from three testing cycles (2000, 2003, and 2006) are currently available; 2009 data will likely be available in 2010.

WHAT IS THE PURPOSE OF PISA?

The PISA assessment probes what students can do with the reading, mathematics, and science that they have learned. It measures students' ability to apply what they have learned to real world situations and to communicate solutions to others. PISA is not tied to curriculum or schooling per se. Rather, it "move[s] beyond the school-based approach towards the use of knowledge in everyday tasks and challenges." Since the emphasis is on application, the developers use the word "literacy" to refer to the primary domains assessed: reading literacy, mathematical literacy, scientific literacy.

The developers believe that success on the PISA test is ultimately related to student motivation, self-confidence in one's own learning ability, and positive learning strategies. Thus PISA's supplementary background questions focus on students' and principals' attitudes. Questions for principals, for instance, include, "Are some kinds of teaching and schools more effective than others?" and "Can schools contribute to improving the futures of students from immigrant or disadvantaged backgrounds?" Students are asked to what extent they agree with such statements as, "I will use science in many ways when I am an adult" and "Science is very relevant to me." These attitudinal questions are embedded (and clearly marked) within the test, but answers to them are not included in their overall score.

WHO TAKES PISA?

The two-hour test is administered to fifteen year-olds who are in grade seven or above. This aims to capture a significant portion of the student population about to leave school, since compulsory education ends in most countries around age fifteen. Schools, rather than individual students, are selected to participate in the test; once selected, all age-appropriate students within the school take the test.

All of OECD's thirty member countries participate, as well as a growing number of other nations; approximately sixty countries participated in the 2006 test.

WHAT DOES THE TEST LOOK LIKE?

The PISA test is comprised of an achievement section, which tests math, reading, and science literacy, and the aforementioned student background questionnaire. Each testing cycle emphasizes one of the three subjects, meaning that a greater proportion of test questions is devoted to that subject; PISA 2000 emphasized reading, 2003 math, 2006 science, 2009 reading, and so on.

The test includes multiple choice, short answer, and extended answer questions. A few questions are generally grouped together around one "stimulus item," such as a reading passage, image, or graph.

HOW IS THE TEST DEVELOPED AND ADMINISTERED?

The test is overseen by the PISA Governing Board (PGB), which includes one voting representative from each OECD member country and one non-voting representative of each non-OECD "participating partner." The governing board chooses a consortium of international contractors to design and implement the test.³⁷

Subject Matter Expert Groups chosen by the PGB design the theoretical framework. The test questions are then designed by

³⁶ PISA 2003 included a trial "problem solving" section.

³⁷ In the 2000, 2003, and 2006 iterations, the consortium was led by the Australian Council for Education Research (ACER); other members included Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe) at Université de Liège in Belgium, the National Institute for Educational Measurement (CITO) in the Netherlands, the National Institute for Educational Research (NIER) in Japan, and Westat and the Educational Testing Service (ETS) in the United States.

members in some of these groups, along with additional individuals in participating countries. All questions undergo further development by contractors and other experts. After much vetting, draft items are piloted with a large sample of students in multiple countries in the relevant age group; the most reliable items are reserved for the question bank.

THE FUTURE OF PISA

For the upcoming testing cycles (2012 and beyond), PISA has three goals: to examine the relationship between the process of education and achievement outcomes; to evaluate how students attain knowledge and skills over time; and to better use and integrate technology—both in how students learn and in how PISA can be adapted for computer-based administration.

Sources:

Adams, Ray and Margaret Wu, eds. 2002. *PISA 2000 technical report*. Paris: OECD. http://www.oecd.org/document/7/0,3343,en_32252351_32236159_33688711_1_1_1_1,00.html

Organisation for Economic Cooperation and Development (OECD). No date. "FAQ: OECD PISA." Paris: OECD. http://www.oecd.org/document/53/0,3343,en_32252351_32235731_38262901_1_1_1_1,00.html

----- 2005. *PISA 2003 technical report*. Paris: OECD. http://www.oecd.org/document/13/0,3343,en_32252351_32236173_35188685_1_1_1_1,00.html

----- 2009. *PISA 2006 technical report*. Paris: OECD. http://www.oecd.org/document/41/0,3343,en_32252351_32236191_42025897_1_1_1_1,00.html

----- *PISA: The OECD Programme for International Student Assessment*. Paris: OECD. <http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf>