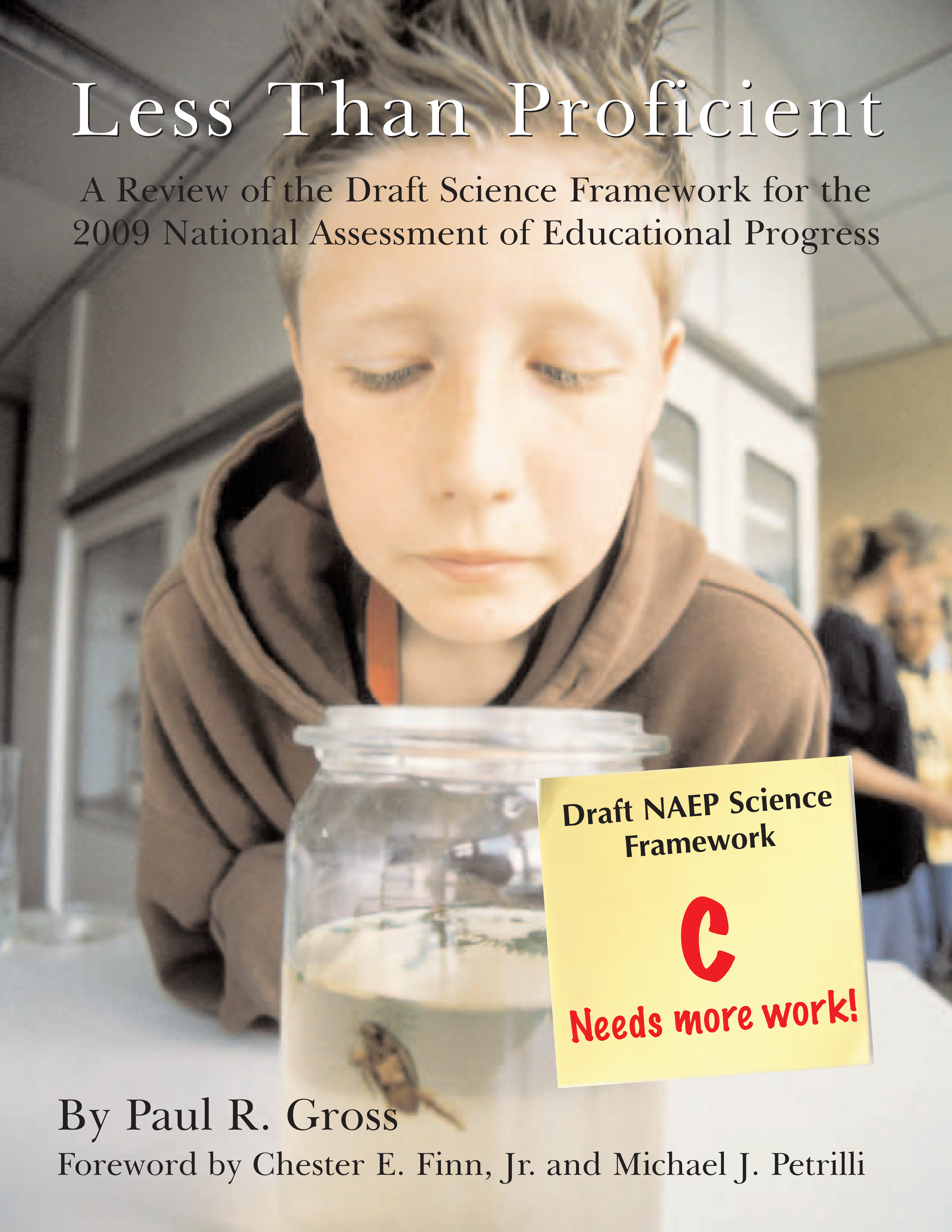
Less Than Proficient

A Review of the Draft Science Framework for the 2009 National Assessment of Educational Progress



By Paul R. Gross

Foreword by Chester E. Finn, Jr. and Michael J. Petrilli

Less Than Proficient

A Review of the Draft Science Framework for the
2009 National Assessment of Educational Progress

By Paul R. Gross

With
Lawrence S. Lerner
Susan Haack
Martha Schwartz
Richard Schwartz

Foreword by Chester E. Finn, Jr. and Michael J. Petrilli

October 24, 2005

Table of Contents

Foreword

A panel convened by the National Academies earlier this month called for "an urgent and wide-ranging effort to strengthen scientific competitiveness," as reported in *The New York Times*. "Decisive action is needed now," declared the scientists. Indeed, ever since the publication and wild success of Thomas Friedman's *The World is Flat*, the nation's policy and business elites have been gripped by concern about the demise of America's longstanding lead in all things scientific and the implications for our economy and the nation's role in global affairs. Much will have to happen for the United States to regain its "national greatness" in the scientific arena, but the most pressing change involves the teaching of science in our schools.

There are grounds for hope. U.S. eighth-graders raised their international standing on the most recent Trends in Mathematics and Science Study. And states are gearing up to implement regular testing in science starting in 2007, as required by the No Child Left Behind Act, which will shed much-needed light on the performance of students, schools, and school systems in this core subject. In connection with the new testing obligation, many states have reworked their academic standards for K-12 science.

Within this context, the National Assessment Governing Board (NAGB) is also making important changes, preparing to launch a new science assessment in 2009 (the National Assessment of Educational Progress, or NAEP). It will replace the NAEP science assessments that have been in use since 1996 and will begin a whole new "trend line" by which America's progress in science will be tracked in grades 4, 8, and 12.

NAEP tests are not "high stakes" exams, but they are enormously influential in shaping America's expectations for what schools should teach and students should learn, and in setting benchmarks by which to gauge "how good is good enough." Many states and school systems model their curricula on NAEP's "frameworks" and strive to align their own tests with NAEP's. So do textbook publishers, teacher educators, and many others.

Creating and implementing a new NAEP test is complicated, time consuming, expensive, and disruptive. NAGB doesn't do it very often and, when it does, close attention must be paid to make sure it's done right. The process begins with construction of a subject-specific "framework." If NAGB's new science framework is crafted properly, it could contribute to setting American science education on a more promising path and giving the nation a jump-start on regaining its scientific leadership. If crafted poorly, however, it could help make today's worrisome situation even more menacing to the country's future.

Our mission here is to appraise the new draft NAEP science framework and to determine whether it is up to snuff. Fortunately, we were already in the middle of a review of state science standards. (Since 1997, the Thomas B. Fordham Foundation has completed several reviews of state standards in English, mathematics, U.S. history, geography, and science. Our most recent review of science standards was published in 1998, with an update published in *The State of State Standards 2000*.) We asked our ace review team on the current science state standards project to examine the NAEP framework, and happily they concurred.

That team is led by Paul R. Gross, one of America's distinguished scientists. Since earning his Ph.D. in zoology from the University of Pennsylvania, he has been professor of biology at five universities and is now University Professor of Life Sciences emeritus at the University of Virginia, where he also served as vice president and provost. He has been dean of graduate studies at the University of Rochester and, for a decade, was president of the Marine Biological Laboratory at Woods Hole, Massachusetts. The winner of prizes, fellowships, and honorary degrees, and a member of many scientific societies, Dr. Gross has authored or co-authored abundant papers, books, and research reports, primarily in biology. Among his recent works is *Creationism's Trojan Horse: The Wedge of Intelligent Design* (Oxford University Press, 2004), written with Barbara C. Forrest.

Paul enjoyed the assistance of four expert reviewers: Lawrence S. Lerner, Martha Schwartz, Richard Schwartz, and Susan Haack (biographical sketches are available in Appendix B). All but one member of the reviewing group are experienced science teachers; one teaches, among other things, philosophy of science. Their combined expertise covers K-12 science, university science education through postdoctoral studies, scientific research, and the management of large research enterprises. Their disciplinary expertise ranges from physics, chemistry, biology, and geology to environmental science, epistemology, and logic.

What follows, then, is a review of the Framework using essentially the same criteria developed for examining state science standards. The review team did make small adjustments in their criteria and judgments, however, recognizing that a NAEP Framework serves a purpose different from a state standards document. The adjustments are described below.

Some readers may disagree with the criteria used by the reviewers (and listed on pages 5 through 8). So be it. But readers (and members of the National Assessment Governing Board) who share our convictions regarding the importance of strong content in the K-12 science curriculum, balanced with an appropriate (but not over-simplified) use of "inquiry-based methods," should find this review compelling. Our basic position is that every child in America should receive a rich and rigorous science education in the primary and secondary grades, one that provides a broad understanding of key scientific concepts and ways of thinking. We reject the trendy notion that children, unaided, can "discover" key scientific concepts. Most of science must be taught if it is to be learned.

In these ways we demur from the "consensus" represented by several professional organizations that have offered national guidelines for science education (the *National Science Education Standards* of the National Research Council and *Benchmarks for Science Literacy* of the American Association for the Advancement of Science [AAAS]). Unfortunately, the NAEP Science Assessment Steering Committee encouraged the Framework's authors to rely upon those very documents as their guide stars. This was a mistake. Just as the National Council for the Teaching of Mathematics (NCTM) is a partisan in the "math wars," so, too, do these organizations represent one pole of the debate over science education and instruction. To follow their guidance is to "take one side" in an important debate rather than to strive for balance.

The importance of that debate is signaled by the fact that several states explicitly rejected the Standards and Benchmarks documents when drafting their own science standards, determining them to be faddish and thin on science content. California, arguably one of the foci of the scientific world, is such a state. Massachusetts is another.

So, with that background in mind, what did our reviewers conclude? Is the framework up to the challenge at hand? In a word, no. The framework received 40 points out of a possible 63, or 62 percent, based on the scoring rubric described within. **This amounts to a grade of "C," or roughly the middle of the pack compared to the states.** Put differently, approximately half of the states faced a challenge similar to that faced by the authors of this draft Framework and produced a better document.

We conclude that the Framework draft is not ready for prime time, and that the

National Assessment Governing Board should make needed repairs before using it as the basis for NAEP assessments. We're well aware that the clock is ticking for the 2009 science assessment, but with some 42 months between now and the first administration of the new assessment, we submit that NAGB can find the few weeks or months that such repairs may consume. It would be a worthy use of time.

In no way are we or our review team suggesting a total overhaul. Among the positive features of the current draft:

- The achievement level descriptions are clear and appropriately differentiated.
- The document itself is clear and well organized, and it contains less jargon than typical state standards documents.
- Its handling of evolution is excellent. It is straightforward, faithful to real science, avoids pseudo-science, and could serve as a model for state standards in this regard.

But much work remains to be done. In particular:

- The expectations expressed in the framework are too modest. We should ask more from our students than did the drafters of this framework.
- When it comes to basic scientific content and students' knowledge thereof, there appears to be neither sufficient breadth *nor* depth.
- Conversely, the draft focuses overmuch on skills and "practices."
- The document contains scientific errors and misleading statements. For example, it declares that "waves are caused by disturbances," when, in fact, waves are disturbances themselves. Though such errors are for the most part minor, given the visibility and influence of a NAEP Framework, its framers should have zero tolerance for errors. (A partial list of errors is included in Appendix A.)
- Mathematical calculation is missing entirely. While perhaps NAGB wants to avoid conflating an assessment of science with an assessment of mathematics, it is impossible to avoid calculations, for the sciences rest heavily on a mathematical base, and rest ever more heavily as the science grows more complex.

Dr. Gross and his colleagues conclude by recommending that NAGB table the draft Framework at its November meeting and give its staff and contractors time to make the following improvements:

1. Aim higher. Delete the instructions (p. 12) that the Specifications focus on the "significant information and knowledge that students should retain…over time, such as ten years after they leave school." That's akin to saying "assess only the residue of today's teaching." It's simply inadequate for tomorrow's America.

2. Rethink all of Chapter Three: "Science Practices." Incorporate important pieces into Chapter Two, including a more sophisticated approach to "Scientific Inquiry." Eliminate redundancies generated by the imaginative (but ultimately unsuccessful) device of "crossing" content and practices.

3. Beef up the content. Include additional examples, especially in areas where the authors intend for students to "go deep." The chemistry content especially needs attention.

4. Incorporate mathematics as needed. Find appropriate places, such as in the fleshing out of Newton's Laws of Motion, where students are expected to use mathematics to demonstrate their understanding of important scientific principles.

5. Give the document a good scrub. Making the corrections listed in Appendix A would be a good start; however, these are just a few examples of errors and misleading statements. A comprehensive solution entails enlisting scientists to review the final Framework and ensure that all content is clear and accurate. (We note that the various committees involved with this draft framework were a lot heavier on science educators than scientists.)

We hope Governing Board members and other readers find this analysis helpful. We urge NAGB to take these and other concerns seriously, to go back to the drawing board, and to spend a bit of time making key edits and needed improvements. After all, states such as California have managed to develop superb science standards. Why should NAEP settle for less? "The nation's report card" is a respected and powerful tool; its authors should be happy with nothing but an "A." This is a solid first effort, but not good enough.

• • •

The Thomas B. Fordham Foundation is a nonprofit organization that conducts research, issues publications, and directs action projects in elementary/secondary education reform at the national level and in Ohio, with a special emphasis on our hometown of Dayton. Further information can be found by surfing to www.edexcellence.net or writing us at 1627 K Street, N.W., Suite 600, Washington, D.C. 20006. The Foundation is neither connected with nor sponsored by Fordham University.

—Chester E. Finn, Jr. and Michael J. Petrilli, October 2005

Introduction

This is an evaluation of the September 30, 2005, draft document, *Science Framework for the 2009 National Assessment of Educational Progress* (Framework), developed for the National Assessment Governing Board.

We are an evaluation group studying the K-12 science standards issued by 49 states and the District of Columbia. Our report on those standards will be released in December 2005. A process very similar to that used in generating that report can be used to evaluate the Framework. Recognizing, however, the difference between a central focus on assessment design, as in the NAEP document, and a focus on curriculum design, as in the state standards, we have made minor adjustments to the criteria and the scoring process.

We list the criteria first and follow that with an account of their application to the reviewing task. We then express the quantitative outcome as a set of average scores for each of 21 criteria, then as the average summed score for all criteria (the "raw" score), and finally as a conversion to an average percentage score ("final score"). On the basis of the final score, a letter grade is assigned the Framework, using the familiar A through F scale. A series of short summary statements responds to questions on which NAGB seeks feedback. In closing, we undertake briefly to justify this response; that is, to convey our appreciation of some qualities of the Framework and our concerns about others, and to make some recommendations.

In the course of this review, we generated detailed comment on the selection and the statements of science content from the base disciplines: physical science, life science, and Earth and space science. There was much discussion of the central design feature of the Framework project: "crossing of science content and science practices to generate performance expectations" (17, 18). Although that feature *is* first a device for the design of assessment items, it is also, inevitably, a potential guide to curriculum standards.

Since the Framework claims to be based on the available national guidelines (the *National Science Education Standards* of the National Research Council and the American Association for the Advancement of Science (AAAS) *Benchmarks for Science Literacy*) but is more recent than those, it is certain to receive close attention from the curriculum-makers of every state. In the interest of brevity, the details of our findings, especially on the science content actually presented in Framework Chapters Two through Four, must be represented by example only. We have taken care, however, to provide representative examples.

Criteria

The criteria are shortened from those used for our 2005 study of the state standards. For the Framework analysis, there are 21 criteria placed under five categories. A few instances of apparent overlap between criteria in different categories may be noted. However, the different categories require different approaches to judgment; hence the overlap does not amount to redundancy.

Group A: Expectations, Purpose, and Audience

1. The expectation is unambiguous that throughout the primary and secondary grades all students will become scientifically literate, at levels appropriate to grade.

2. The standards can be used in designing effective assessments of student learning, theoretical and practical, appropriate to grade.

3. The presentation is as free as possible of jargon; it is lucid and comprehensible to all its audiences: educators, subject matter experts, policy makers and legislators, parents, and the general public.

4. The standards call for student written work in good English and, where appropriate, in suitable mathematical language. They require student oral presentations that are clear, logical, and appropriate to grade.

Group B: Organization

1. Standards are referred to grade or to clusters of no more than four grades.

2. They are grouped in categories or themes that reflect the fundamental theoretical structures in modern science. Examples: Newtonian dynamics; conservation of mass and energy; cosmological evolution; plate tectonics; cells and organisms, inheritance, populations and ecosystems, organic evolution.

3. Instruction in the themes is devoted at each grade level to developing skills of observation and data gathering; to the planning, recording, and interpretation of observations; and ultimately to the design of experiments.

Group C: Science Content and Approach

1.The standards provide explicitly for substantial laboratory and (as appropriate) field experience. Replication of classical experiments is encouraged. The importance of empirical evidence and of sound criteria for the acceptance of data is emphasized.

2. Unambiguous terminology and rigorous definition are stressed. Such terms as energy, mass, valence, pH, genotype, natural selection, cell, metabolism, continental drift, magnetic reversal, and cosmic background radiation are defined as carefully as possible for the grade level in which they are introduced.

3. At appropriate grade levels, data analysis, experimental error, reliability, and the practices needed to optimize the quality of raw information are taken up—as subject matter.

4. The standards call for mastery of tabular and graphical techniques for analysis and reporting, with increasing sophistication as grade succeeds grade.

5. The continuing interplay of data and theory, and well-justified modifications of theory, are stressed at all grade levels, in a manner commensurate with student maturity. Important conceptual shifts and innovations in the history of science are an element of the curriculum.

6. The primary curriculum content is an adequately representative set of basic principles, explicit or contained within science themes. Examples (only) of basics: In physics, Newton's laws of motion, conservation laws, the macroscopic/ microscopic nexus; in astronomy, evolution of the universe and the structure of its parts (including the solar system); in geology, planetary structure, plate tectonics; in chemistry, mass and energy conservation, atomic structure and the nature of the chemical bond; in biology, cells, organisms, ecosystems, biochemical unities, history of life and evolution.

7. These principles are first introduced via facts and simple examples; they emerge as themes and theory in higher grades. Increasing ability of students to grasp generalizations and abstractions is taken into account. An adequate factual knowledge base, laid down in the early grades, is deepened systematically by means of increasingly refined theory.

8. The standards emphasize recognition of good inquiry as well as some of the distinctive methodologies of natural science, but they do not oversimplify these as "*the* scientific method." Common features of every kind of competent inquiry, including good science as well as distinctions among different disciplines, are made clear.

9. The standards provide for careful definition of technology and do not confuse it, or its social consequences, with the content of science. They do address relationships between science and technology and the way that science has shaped the modern world.

Group D: Quality

1. The standards are demanding as to science-disciplinary content; their expectations are neither so broad as to be vague nor so narrow as to be trivial. They are neither mere prosy encouragements nor simple lists of things to be memorized.

2. They cover many of the basic understandings of physical reality as the scientific community recognizes them, but the document makes no effort to be encyclopedic.

3. The standards, taken as a whole, define a core scientific literacy for all students in all schools of the state. At the same time, however, they are sufficiently challenging to ensure that students who achieve proficiency by the final year will be ready for college work.

Group E: Seriousness

1. Nowhere do the standards offer or encourage—as though they were science—pseudo-scientific or discredited proposals such as medical doctrines not based on objective evidence, vaguely defined "energy fields," "auras," folk-cosmologies and mythologies, creationist or neo-creationist anti-evolutionism disguised as "critical thinking," UFO visits, astrology, or divination.

2. Nowhere do the standards suggest or imply that basic scientific principles are race-,

ethnic-, or gender-specific; nor do they distort the history of science in an effort to inculcate social or political doctrines.

Application of the Criteria

The degree to which the Framework meets the requirement of a criterion was measured on a four-point scale:

0—The requirement is not met, or its treatment is useless
1—The requirement is addressed, but incompletely, erratically, or inconsistently
2—The requirement is addressed adequately but with no distinction
3—The requirement of the criterion is met, and in a thoughtful manner

The Framework was read by all members of the reviewing group, each reader giving special but not exclusive attention to subject matter (science content) within his or her professional expertise. Throughout the study, we exchanged views on Framework features. Our assigned final grade is a consensus.

The maximum number of points possible is 63. The aggregate score for each reader was the sum of his or her criterion scores, expressed as a percentage. The final score is the average of those percentage scores. This average was the primary datum in the assignment of a letter grade. The device of averaging should not be interpreted to mean that the quality of science content was the same for the three base disciplines. The Framework's selections of science content from among them are not equally meritorious. We discuss this below.

Table 1

Average scores by criterion and final raw and percentage scores.

Criterion Number	Average Score (Range: 0-3)
A1	1.50
A2	1.25
A3	1.75
A4	1.50
B1	2.00
B2	2.50
B3	2.25
C1	1.75
C2	1.00
C3	1.00
C4	1.75
C5	2.00
C6	1.75
C7	1.75
C8	2.25
C9	2.25
D1	1.75
D2	1.50
D3	1.50
E1	3.00
E2	3.00
Average(Raw) Total Score	39
Final Score (Percent of Maximum)	62

Number and Letter Grade

Table 1 shows, by criterion, the readers' average evaluation, together with the average combined raw score and the derived percentage (final) score. The latter is 62 percent, which, in the schema of letter grades we have used for the state standards, yields the grade C. (The A range is 85-100%; B is 70-84%; C is 55-69%; D is 45-54%; and F is 44% or below.) We intend this "C" grade to signify passing and adequate, but not an honors performance. The Framework is an interesting start, but there is much work to be done if it is to achieve its potential usefulness.

Responses to NAGB Questions

1. Does the science content described in the draft represent important information and reasonable expectations for student achievement? Is there an appropriate balance among different fields of science at grades 4, 8, and 12?

Yes, the science content, so far as it goes, *represents* important information. But it does not go far enough. Perhaps in a later draft, the promised "unpacking" of content touched upon in this one, together with the as yet unavailable *Specifications* document, will add sufficient content to mitigate our concern. Nevertheless, in this draft, the indicated science content, all of which *is* important, is generally too thin. Emphasis for the chosen main fields of science is roughly "balanced."

2. Are science skills and practices treated satisfactorily and given appropriate emphasis? Should technological design be part of the assessment?

Science skills and "practices" are given too much weight relative to the weight of science content that a reader of the Overview (Chapter 1) would expect. The key distinction between *principles* and *practices*, although heavily labored in the document, is fatally soft. Many of the putative assessments of "practices" will simply test content knowledge. That is fine with us, so long as the assessors understand what they are doing.

3. Are the preliminary achievement level descriptions reasonable, clearly expressed, and differentiated appropriately?

The generic achievement level definitions (p. 128, for example) are clearly stated and well differentiated, provided it be understood that "partial mastery," as employed in the definition of the lowest level ("basic"), includes nothing that is described by the word *mastery* as it is usually employed in connection with science content.

4. Is there an appropriate mix of item types—multiple-choice, short answer, and long constructed response? Should interactive computer-delivered tasks and hands-on experiments be part of the assessment as recommended?

It is difficult to answer this question; there are too few samples provided in the draft. We assume that in the actual assessment, items of the first three kinds will be used in roughly equal numbers. That would be appropriate, given the increasing potential difficulties, along the sequence of those types, in evaluating the test-taker's answer. It is true, however, that good multiple-choice questions can sometimes be as difficult to design as the other types.

Based on the few examples provided in the draft, we are skeptical about the computer-based and hands-on items. Correct or good answers to computer-delivered tests are as much influenced by the test-taker's comfort with computing and familiarity with the relevant software as by content knowledge or practice skill. Hands-on assessments are notoriously hard to design and, even more troubling, to grade.

5. Does the draft framework provide clear information about the content and format of the NAEP science assessment? Are the organization and format of the document appropriate for the intended audiences, including the general public, teachers, policy-makers, curriculum specialists, and scholars?

The draft framework is well organized and clearly written. Inevitably, it contains jargon that will be mysterious to lay readers, but the density of jargon is lower than in most state science standards documents. Our concerns are not with the Framework as a public document, but with some—by no means all—of the underlying assumptions about appropriate science "principles" and "practices" that guided its preparation.

Plaudits

The document is carefully organized. The organization is nicely outlined and explained in the executive summary. This Framework is well-written, and such jargon as does appear is routinely professional. The body of the document follows with precision the plan outlined in the Table of Contents and the Executive Summary, so that—unlike a great many current state standards documents—this Framework is easily navigable.

In a number of places, the statements of Scientific Principle (content) are clear and correct. Much thought has gone into their formulation, especially in the examples of "unpacking." Some of the writers know the science they write about. It is reasonable to expect that this level of care and expertise will be sustained as the range of content and of its dependent assessment items expands.

As is appropriate for an assessment framework, the Framework examines at considerable length the issues of item construction and item type. It is frank enough, in these sections, about anticipated difficulties of design and use, especially with hands-on and computer-generated tasks. Finally, although "technological design" is one of the four "science practices" upon which the basic plan of assessment is built, the document reports forthrightly a disagreement about its appropriateness. We agree with those of the project staff who question it.

Finally, there is a very sound and especially succinct statement on "Evolution and diversity of living systems" (p. 45).

Concerns

1. Low Expectations

This concern is greatest regarding the Overview (Chapter One), which conveys the steering committee guidelines for the planning of this Framework (pp. 11-12). Explaining the important connections among the Framework, the (as yet unavailable) *Specifications*, and the actual assessments to come, the document describes as its target, "the significant information and knowledge that students should retain (e.g., big ideas, fundamental understandings) over time, such as ten years after they leave school." This is unfortunate.

To begin with, one needs to know what involvement with science of the tested population is to be expected a decade after it has left school. On average, so far, that involvement is very small. Science knowledge retained by high school graduates, as a population, is minimal. Only the small subpopulation attending college and majoring in a science or engineering has real interest or involvement after ten years. But in its opening pages (e.g., p. 4) the draft expresses lofty goals in this regard, not the least of which is "to provide young people who choose to pursue careers in science and technology with a strong foundation for their post-secondary study and work experience."

What must be "provided" is a far stronger base of science knowledge than is possessed by today's average high school graduate plus-ten. Thus a first concern is that the knowledge expected for "proficiency" should be both broader and deeper then that of today's "man in the street." More serious, however, is the idea implicit in much of the Framework's discussion—that only what can be retained after a decade needs to be included as core content, and assessed. This idea is false.

The universal experience of people working in science is that re-acquisition of knowledge one has had in the past, but has now forgotten, happens and is even easy—*if* it was once learned properly. Acquisition of knowledge *de novo* is much harder.

For example: few physicians have any need for mathematical techniques such as the handling of exponentials in calculus. Should they engage, however, in a refresher course in modern pharmacology, such handling will be indispensable, and the new knowledge of pharmacodynamics will be important in their practices. *Recovering* needed skills of algebra and calculus is usually straightforward; acquiring them *ab initio* is difficult or impossible for a busy adult who has never done such mathematics.

What pedagogues have always claimed about this phenomenon is self-serving but true: The absolute content of a school subject is not necessarily retained, but the early discipline of learning it is essential for later re-acquisition. What we see in this draft, however, is a conscious effort to reduce science content to "significant" knowledge. The Framework's notion of "significant" knowledge is far too modest for a student body whose scientific literacy is to be raised above that of the current population.

It is possible that the promised "unpacking" of science content statements, to be found in the forthcoming *Specifications*, will lead to far more searching and interesting items than are shown in this draft, and that they will make our concern gratuitous. It is possible that the content-minimizing strategies of the national standards guides, to which the writers of this Framework credit inspiration, will be fleshed out in that unpacking so that the claim of "depth instead of breadth" will be justified. We certainly hope so, but so far we see neither breadth nor sufficient depth.

2. Indistinct Distinctions

A single key device for delineating performance expectations, hence for designing assessment items, dictates the structure of this Framework. It is "crossing content and practices" (pp. 17-19 ff.). "Crossing" means the creation of a large matrix, one of whose principal dimensions is the set of three science content areas (Physical, Life, Earth and Space Sciences). Another is a set of four "Science Practices" to which at least 30 percent (but probably a larger proportion) of the required science learning refers. Each cell, each intersection of every assessable science content statement with each of four science "Practices," becomes a performance expectation.

This is an imaginative device, offering a schematic means of boosting the importance of "Inquiry" as science knowledge. But it is not a simple device. The science *Practices* are the realization, in this plan, of the "Science as Inquiry" pedagogy adopted in the national science standards models. The four Science Practices are:

- Identifying Science Principles,
- Using Science Principles,
- Conducting Scientific Inquiry,
- Employing Technological Design.

However, as is admitted repeatedly in the text (e.g., p. 66, first paragraph), these categories are not distinct. The best description is "fuzzy." The first and most important of the four is not *at all* distinct from the "Science Content" dimension. One reader reported, somewhat wryly, " 'Identifying Science Principles'…means, so far as I can tell, 'students can remember and state some basic scientific ideas.'…But the idea that this, which sensible science teachers have been doing forever, is new, and needed the 'extraordinary effort…' etc., is just sad."

In short: The first Science Principle, "Identifying…," is just *knowing* the science content. Efforts to "cross" the one with the other are not likely to yield conceptual clarity for authors of assessment items.

The second Science principle, "Using…," means that students can apply some of those scientific ideas ("content") to some simple examples of natural phenomena. This isn't much, to begin with, since the emphasis is clearly to be on "simple." More disconcerting is the illustration provided for such an application (p. 69). This is not at all simple. It can have a number of equally good answers, depending upon specific information (e.g., the feeding habits of snakes and hawks), information that is not provided. The potential problems of scoring answers to such questions are fearsome. We are not reassured that an adequate supply of good, unambiguous "Using…" items will be available as needed.

As to Conducting Scientific Inquiry, the Framework asserts that it is "a complex and time-intensive process that is iterative rather than linear." (p. 71) We don't know what this distinction is supposed to mean. Perhaps it means that science is continuously self-correcting. If so, then it is a good statement.

The remainder of this paragraph is self-congratulatory. But asserting that scientific inquiry (and no other inquiry?) requires "logical" reasoning and "relevant" data does not advance the understanding of readers eager to create "relevant" and non-trivial assessments of Scientific Inquiry. From context (pp. 72 ff.) it is evident that "Conducting Scientific Inquiry" is well-meant. It includes, notably, recognition of patterns in data and the fitting of those patterns to theoretical models of the phenomena under study. But these good intentions are vitiated by the examples provided, whose language is kept artificially simple, and whose content is at too low a level of sophistication, especially for grade 12.

The basic design argument for this Framework, in short, is ingenious, but it's flawed by the attempt to make distinctions where distinctions are difficult or impossible to make.

3. Depth versus Breadth?

About the decision to limit breadth of science content coverage in favor of depth there can be friendly argument. There is certainly insufficient research of highest quality to ensure that the decision to do so is right. Still, it is the decision made by the *National Standards* and by *Benchmarks*, together with follower documents such as the National

Research Council's *Inquiry* (National Academy Press, 2000). It is possible that a K-12 science education devoted to just a few samples of science content, but at significant depth, will be a good education.

In fact the idea is ancient, and not just in natural science. It is for example the basis of every "Great Books" curriculum. Everything depends, however, upon the actual depths to be plumbed. We see no evidence that more than the current, ordinary depths of science understanding will be probed by assessment items to be written under these guidelines. Perhaps this judgment is mistaken; perhaps those forthcoming *Specifications* will so enrich the truncated content of the three main science disciplines as to constitute a true search for depth of understanding.

4. Calculate?

In the Framework as it stands now, there seems to be active avoidance of quantitative argument (in the sense of dependence upon actual mathematical work). The word "calculate" is vanishingly rare. If this is intentional and not just an accident of the selection of samples for the draft, it will be unfortunate. Very little modern science can be done without mathematical calculation. Of course, the assessments are not supposed to be concerned with mathematics per se. But natural science depends on mathematics. Some of its most important "Principles" cannot be understood or "Used" without mathematics. We hope that in fleshing out such principles as Newton's Laws of Motion, or the dependence of population size on variables such as reproductive rate and predation, efforts will be made to include appropriate calculations.

5. Quality Differences by Discipline

The treatment of content and practices referred to the three main disciplines is uneven. Physics does best. A reviewer reports: "The physics content…is generally well organized and mostly correct, but some areas seem to have been shortchanged…there is very little about geometric optics, almost nothing about physical optics, little about electromagnetic induction (Faraday's and Ampère's laws) or electrical devices, nothing about the laws of thermodynamics."

Chemistry, among the physical sciences, is notably shortchanged, and this is also the case for many of the state science standards documents—even in some of the best. Why this should be so for chemistry and less so for physics is a mystery.

Like all the others, the life sciences are thin on breadth of important subject matter, and they are not very deep either. Nevertheless, what is offered is mostly correct. The essential ideas of evolution are well represented and well stated. This does not mean that there are no errors.

The writing is workmanlike in earth sciences. Though not particularly elegant, it is not plagued with errors of the kind encountered in some state standards. It suffers from the usual neglect of the fundamentals of the solid earth, mentioning minerals only in a list of things. It relegates important processes like earthquakes, vulcanism, and erosion to a thematic gloss such as "slow" or "fast," or vaguely to plate tectonics. The lack of physical science underpinning for much of it—concepts such as air pressure, specific heat, the gas laws—renders the earth sciences treatment shallow. Perhaps the forthcoming "unpackings" will eliminate these problems.

Recommendations

The National Assessment Governing Board should vote to table the Framework at its November meeting and give its staff and contractors time to make the following improvements:

1. Aim higher. Delete the instructions (p. 12) that the Specifications focus on the "significant information and knowledge that students should retain…over time, such as ten years after they leave school." That's akin to saying "assess only the residue of today's teaching." It's simply inadequate for tomorrow's America.

2. Rethink all of Chapter Three: "Science Practices." Incorporate important pieces into Chapter Two, including a more sophisticated approach to "Scientific Inquiry." Eliminate redundancies generated by the imaginative (but ultimately unsuccessful) device of "crossing" content and practices.

3. Beef up the content. Include additional examples, especially in areas where the authors intend for students to "go deep." The chemistry content especially needs attention.

4. Incorporate mathematics as needed. Find appropriate places, such as in the fleshing out of Newton's Laws of Motion, where students are expected to use mathematics to demonstrate their understanding of important scientific principles.

5. Give the document a good scrub. Making the corrections listed in Appendix A would be a good start; however, these are just a few examples of errors and misleading statements. A comprehensive solution would involve hiring scientists to review the final Framework and ensure that all content is clear and accurate. (We note that the various committees involved with this draft framework were a lot heavier on science educators than scientists.)

Appendix A: Science Errors and Misleading Statements

Most of the mistakes discovered in our reading were minor, and these were to some extent offset by correct statements of the same material elsewhere. Nevertheless, there are some outright errors and too many misleading statements in this limited set of content items. A few samples follow to show the nature of the slips.

Physical Sciences

1. Table 2, (8^{th} Grade): "If more than one force acts on an object along a straight line, then the forces will reinforce or cancel…" This is misleading. "Cancel" usually means "negate." A better statement: "…forces will reinforce or oppose…" (The misleading statement is also in the *National Standards*.)

2. Table 3, Unpacking: "Waves are caused by disturbances. Some of the energy of these disturbances is transmitted by the wave." But waves themselves are disturbances. The most fundamental definition of a wave is that it is a disturbance in a medium.

3. Page 31: The textbox (Grade 12) says that plasma is not included in the Framework since it is foreign to the experience of students. How about the fluorescent lights overhead, in the classroom?

4. Page 38, P12.12: The term "ions" must be included alongside "atoms" and "molecules" to allow for ionic crystals, which are, of course, ubiquitous.

5. P12.2: "Forces that hold all the particles in the nucleus together are stronger than the electrical forces between the positively charged protons." No. If the nucleus is at equilibrium, the (repulsive) electrical forces must be equal to the attractive strong forces within it. The distance dependence of the strong nuclear force is, however, stronger than that of the electrostatic force.

6. On Gravity, a *Using…* item: Multiple-choice question. "Is there gravity in space? Which of the following gives the best response to this question?" There are four choices, with Key = c: a. No. You can see the astronauts float around weightless in their cabin. B. No. There's no air in space, so how can there be gravity there? c. Yes. There must be gravity since planets keep circling the sun. d. Some. The moon has one-sixth as much gravity as Earth, so we know there's some gravity in space.

The question is badly phrased and several of the possible answers are poorly chosen. In near-Earth trajectories it is Earth gravity and not Sun gravity that dominates. In an Earth-Moon mission, things are more complicated. For an extra-solar system mission, solar gravity may be negligible. Yes, there is a gravitational field ("gravity") everywhere in the universe, but this question is badly suited to the issue.

7. Page 7: Hands-on Performance Tasks. We have serious reservations about the requirements for hands-on performance and interactive computer tasks. The chance for failure of the task delivery system is high, as is well-known by those who design laboratory exercises in ordinary courses.

8. Page 88: Having students determine the boiling point of unknown liquids can be extremely dangerous.

Earth and Space Science

1. Page 133, Advanced Earth and Space Science suggestion: students to design an experiment to measure "the changes in temperature when water evaporates or condenses." There is no temperature change during a change of state.

2. Page 31: "Unique Properties of Water" ignores specific heat, a serious mistake.

3. Page 52: Contrary to the implication, the "story in the rocks" is not outdated. Crucial evidence for the historical science of geology still comes from the close study of rocks.

4. Page 54: Earthquakes and volcanoes are called "changes" rather than processes, and it is claimed that these two can be observed on human time scales while mountain building and plate movements cannot. But mountain building does occur abruptly during some earthquakes and eruptions. Plate movement can be measured after some big earthquakes.

5. Page 80: "Identify wind as the movement of air from higher to lower temperature regions." What is surely meant here is pressure, not temperature.

Life Sciences

1. On page 41: Paragraph 3 speaks vaguely of levels of understanding biological systems and of interactions within and across levels. Then, unexpectedly, "[U]nderstanding how populations of organisms change over time is greatly facilitated by understanding the changes that occur in DNA molecules." In principle this is perfectly correct. In practice it too is vague. *What* changes in DNA molecules? *How* to "understand" such changes; and what does it all have to do with "interactions" among various levels of organization?

2. p. 42: "Cellular processes are carried out by molecules, particularly proteins." Why *particularly* proteins? Why not nucleic acids? Why not phospholipids, or water, or, for that matter, ions?

3. p. 43, Clarification: "Food." "Pollen" is not a plant food in the sense of this paragraph, nor is "pollen" "synthesized" in the chemical sense used elsewhere in the textbox.

4. p. 45, paragraph 4: This makes it sound as though the only function of DNA is to encode proteins. Not so: Some DNA with important functions does not encode proteins, but rather regulates the function of other DNA, whether or not protein-coding.

5. p. 47, Grade 8 content statement: "Following fertilization, cell division produces a small cluster of cells that then differentiate." "Small cluster" is rather misleading. Cleavages in embryos do produce clusters of cells, but they are "small" only in diameter. Before any kind of differentiation happens in most (not quite all) embryos, the "small cluster" of cells contains thousands or millions, generally *before* any "differentiation" takes place. ("Determination," a commitment to differentiate, is another matter entirely.)

6. p. 50, Grade 8: The word "species" is used here before it is defined rigorously. A definition is crucial to this content statement.

7. p. 50, Grade 12: "Each gene carries a single unit of information." This is vague, and as a generalization, incorrect. Many genes undergo alternate splicing and can thus encode different proteins; others encode non-translatable RNAs, which have multiple functions, and therefore these genes bear more than one kind of "information." The impulse to use the word "information" in the context of genetics is almost irresistible. The impulse is sometimes sound; but here, and at this level, it is not.

8. p. 51, Grade 4. Even here, at this early stage, "differences" is too vague. The real point is that within any population there is, always and necessarily, genetic variation—a vast latency of response to changing environment.

Appendix B:
Biographical Sketches of Author and Reviewers

Paul R. Gross is University Professor of Life Sciences emeritus of the University of Virginia and a Fellow of the American Academy of Arts and Sciences. He is a former vice president and provost of the university. He served as director and president of the Marine Biological Laboratory, Woods Hole, Massachusetts, from 1978 to 1988 and was professor of biology at the University of Rochester and at MIT. In addition to many papers and monographs in cell and molecular biology, and on animal development, he has published widely on science in society. His latest book is *Creationism's Trojan Horse* (Oxford University press, 2004), written with Barbara C. Forrest.

Susan Haack was educated at Oxford and Cambridge. Formerly a Fellow of New Hall, Cambridge, and then lecturer, reader, and professor of philosophy at the University of Warwick, U.K., she has taught since 1990 at the University of Miami, where she is Cooper Senior Scholar in Arts and Sciences, professor of philosophy, and professor of law. Her books include, among others, *Deviant Logic*; *Philosophy of Logics*; *Evidence and Inquiry: Towards Reconstruction in Epistemology*; and *Defending Science—Within Reason: Between Scientism and Cynicism*. She has published numerous articles in professional philosophy journals, in scientific publications, and in such general-interest magazines as *Skeptical Inquirer* and *Free Inquiry*, and in literary magazines, including *The Times Literary Supplement.*

Lawrence S. Lerner is Professor Emeritus in the College of Natural Sciences and Mathematics at California State University, Long Beach, where he won a number of outstanding teacher awards. He is the author of two university-level physics textbooks, co-translator and editor of a seminal work by Giordano Bruno, and author of about 150 book chapters, review articles, journal articles, and patents on condensed-matter physics, the history of science, science and religion, science and society, and science education.

Martha Schwartz has taught science and elementary mathematics for students from seventh grade through early graduate school. She is also experienced in teacher training and teacher professional development. She holds a B.S. in mathematics from Arizona State University, a teaching credential from UCLA, a Master's degree in geology from California State University Long Beach, and a Ph.D. in geophysics from the University of Southern California. She is a member of the Assessment Review Panel in science for the state of California and has worked on school improvement, standards, and testing for a variety of organizations.

Richard Schwartz holds a B.S. degree in chemistry from Arizona State University, a teaching credential from UCLA, and a Master's degree in environmental science from California State University, Dominguez Hills. He taught secondary science for 34 years, the last 32 at Torrance High School in Torrance, California. He is a former member of the California Curriculum Commission and a 1995 recipient of the American Chemical Society's regional award in chemistry teaching. Presently retired from teaching, he helps manage the geochemistry laboratory at the University of Southern California.