Achieve, Inc.

Measuring Up

A Report on Science Standards and Assessments for

NEW JERSEY

ACHIEVE'S BENCHMARKING INITIATIVE

Achieve, Inc.

Achieve is an independent, bipartisan, nonprofit organization created by governors and corporate leaders to help states and the private sector raise standards and performance in America's schools. Founded at the 1996 National Education Summit, Achieve has sponsored two additional Summits in 1999 and 2001.

Achieve helps states raise academic standards, measure performance against those standards, establish clear accountability for results and strengthen public confidence in our education system. To do this, we:

- help states **benchmark** their standards, assessments and accountability systems against the best in the country and the world;
- provide sustained **public leadership** and advocacy for the movement to raise standards and improve student performance;
- build **partnerships** that allow states to work together to improve teaching and learning and raise student achievement; and
- serve as a **national clearinghouse** on education standards and school reform.

EXECUTIVE SUMMARY

Governors and business leaders created Achieve, Inc., after the 1996 National Education Summit to help states raise academic standards, improve student achievement and build support for public education. As a central part of its mission, Achieve provides states with candid appraisals of the quality of their academic standards, assessments, accountability systems and other policies to promote high academic achievement. To date, Achieve has prepared 16 such independent reviews for states.

Last year, at the request of the State Education Department of New Jersey and in conjunction with the Business Coalition for Education Excellence, Achieve conducted an evaluation of New Jersey's *Core Curriculum Content Standards for Science* (1996) and the alignment of its grade 4 and grade 8 science tests against those standards. This review of science standards and tests followed similar reviews of New Jersey's English language arts and mathematics standards and tests in 2000.

Because science standards were slated for revision in summer 2001, Achieve submitted a preliminary set of findings regarding the quality of the standards. This allowed New Jersey to have the benefit of an external review prior to revising the standards.

In December 2001, Achieve presented a final, detailed technical report to the state covering both the standards and tests. This report remains confidential because it references secure test items. This executive summary captures Achieve's major findings and recommendations, as well as appraisals of New Jersey's responses to them.

The report answers two vital questions for policymakers:

- How do New Jersey's science standards compare with exemplary national and international standards?

- How well do New Jersey's tests in science assess the knowledge and skills described in the standards? How challenging are the tests?

STANDARDS BENCHMARKING

Achieve believes that benchmarking standards against national and international models of excellence provides states with meaningful comparisons and specific recommendations for improvement, rather than one group's opinion of the nature and content of "good" standards. Consequently, Achieve's analysis of standards involves a detailed, comparative analysis of a state's standards against exemplary standards by nationally recognized standards and assessment experts in their fields.

Achieve's expert reviewers compared New Jersey's Core Curriculum Standards to the National Science Education Standards and to those of Japan, one of the highest-achieving nations in the

world in science. A summary of their findings and recommendations, as well as New Jersey's response as evidenced in the 2002 edition of the science standards, follows.

- ✓ **While New Jersey's science standards are generally strong, some indicators should be clarified or described with greater precision.**

By and large, the standards and performance indicators are organized in a logical, functional way; are clearly expressed; and are accessible to the general public. In addition, the majority of the performance indicators are measurable or verifiable. Their revision involved refinement, not reinvention.

Achieve recommended that, in refining the standards, the state should strive for greater clarity, specificity and measurability (in those particular instances where indicators focus on activities or the process of learning, rather than the results). Also important was achieving closer alignment with the test specifications, which in some cases detail concepts and skills not evident in the standards.

New Jersey's Response

In revising the science standards, New Jersey improved their clarity by reducing the number of standards from 12 to 10, eliminating redundancies and dividing the standards into subcategories, thereby clearly signaling what the emphasis in teaching and learning should be. Life Science, for example, now is divided into five areas: Flow of Matter and Energy, Structure and Function, Diversity and Similarity, Evolution of Life, and Heredity. Another benefit of this approach is that the subcategories can be lined up from one grade to the next so that progressive development in knowledge and skills across grade levels becomes readily apparent. In contrast to the 1996 edition of the science standards, the progress indicators also are more explicit, generally honing in on what students should know and be able to do.

- ✓ **The science standards present a rich set of important concepts and skills.**

New Jersey has taken care to select fundamental, important content, generally avoiding the inclusion of non-academic or low-level factual standards and indicators. An unusual and praiseworthy feature is the inclusion of a standard that makes the important relationship between math and science explicit. By setting a goal of scientific literacy for every student and attempting to balance content knowledge with both an understanding of the nature and processes of science and workplace skills, New Jersey has placed a premium on the development of critical thinking.

In general, however, the content standards were found to be stronger than the process standards. The two sets of standards were not linked in a substantive way. Reviewers recommended that the thinking skills be more fully described, with attention being given to distinguishing observation from inference, making predictions, identifying sources of bias in data, recognizing that findings may have more than one reasonable interpretation, identifying flaws in the way control groups are constructed, and explaining the difference between cause and effect and correlation. They also recommended that New Jersey link the process and content standards, adding concrete,

reciprocal references to make the connections explicit. Another approach would be to anchor the content progress indicators with examples of how students might demonstrate conceptual understanding of inquiry and the nature of science.

New Jersey's Response

The latest edition of the science standards does not reciprocally cross-reference science process and workplace readiness skills with the content standards to the extent desired. Notable exceptions are Standard 5.3, which specifies multiple, routine applications of math to science, and Standard 5.4, which includes a single reference to workplace readiness in terms of technology. However, New Jersey still could make the connections explicit when it updates both its test specifications and its curriculum framework to align with its revised standards. By developing thoughtful, varied and rich exemplars of student investigations for teachers to use or adapt, New Jersey would more clearly demonstrate the level of rigor it expects. These exemplars could provide controlled tests; field studies; design; and secondary, non-experimental research — all important types of investigations that students should experience in the course of their science education. These exemplars have the potential to unpack the critical thinking and workplace readiness skills that New Jersey obviously values.

- ✓ **The science standards generally lay out a progression of knowledge and skill from grade span to grade span.**

For students to comprehend complex science concepts, they generally need to revisit them over time, moving from observing and describing phenomena to conducting simple investigations to manipulating abstract concepts. Quantitative understanding is built on a strong foundation of qualitative experience. The intent of the cumulative progress indicators is to describe what students should know and be able to do from one grade span to the next in such a way that teaching and learning in the elementary grades provides a firm foundation for more advanced treatment of concepts and principles in middle and high school.

Achieve reviewers found the spiraling of knowledge and skill to be uneven and an area that would benefit from increased attention. To support tighter progression and to assist teachers in deciding on the scope and sequence of learning activities, reviewers recommended that New Jersey consider developing progress indicators for spans of two grade levels rather than four.

New Jersey's Response

The Core Curriculum Content Standards for Science now spell out the content and skills students are expected to acquire by the end of grades 2, 4, 6, 8 and 12 for standards 5.4 through 5.10. Only those standards concerned with Scientific Processes, Science and Society, and Mathematical Applications continue to be described for longer grade intervals, i.e., grades 4, 8 and 12.

✓ **Some essential content was overlooked.**

Chemistry proved to be relatively weak in New Jersey's standards. Also, human biology, a major content area in the Third International Mathematics and Science Study, Japan and the National Science Education Standards, was not addressed. The treatment of geology was inadequate due to the absence of such key ideas as the interrelationship of the geosphere, hydrosphere and atmosphere; the role of the sun in weather systems; and the role of gravity in the universe. Reviewers cautioned that New Jersey's response to essential missing content should be a process of substitution, not addition. For example, two standards, 5.6 and 5.7, were found to be redundant, describing similar content. Combining these standards would provide an opportunity to strengthen the treatment given to critical topics in chemistry, human biology and geology, and also would decrease the likelihood that the content covered in these two standards would end up being overemphasized in teaching and testing.

New Jersey's Response

In revising its science standards, New Jersey also made substantive advances in selecting and laying out significant content in Life Science, Chemistry, Physics, Earth Science, Astronomy and Environmental Studies. Essential areas of study that reviewers had marked as missing now are included. The result is a much stronger document — one that will be invaluable in developing and supporting aligned curriculum and professional development.

✓ **The rigor of New Jersey's science standards does not always match that of the benchmark documents.**

While New Jersey's science standards were clearly influenced by national reform efforts, reviewers found the standards did not match the rigor of Japan's. Depth of understanding is sometimes sacrificed for breadth of content coverage. Balancing depth and breadth is a formidable challenge in a field that is composed of multiple disciplines — physics, chemistry, biology, geology and astronomy — each with its own set of complex research strategies and instrumentation. Nonetheless, certain realities are inescapable: Scientific knowledge is proliferating at an unprecedented rate and, at the same time, advances in information technology are making vast stores of knowledge available and accessible. Today's students will be best served by developing a deep understanding of fundamental concepts on which to build further knowledge and by developing the abilities to analyze and synthesize.

New Jersey's Response

Comparison of the 1996 and 2002 standards indicates that New Jersey has raised the rigor of its standards in a number of important ways. Previously, the treatment of the major disciplines had been uneven, but the revised standards focus on significant content, the "big ideas" in each domain that are the foundation of scientific literacy. In addition, by explicitly linking specific concepts and skills in each of the four content strands in math to corresponding concepts and skills in science, the state has succeeded in making a conceptually powerful standard even more robust. An added benefit is that instructional attention to the intrinsic relationship between math

and science likely will deepen the student's grasp of concepts in math that they often find abstract. Finally, reviewers found it difficult to judge the level of rigor expected by the state in the 1996 edition of the standards because the cumulative progress indicators often lacked specificity. In revising its standards, New Jersey made a concerted effort to address this weakness and can continue to make the expected level of rigor explicit by providing detailed explanations and concrete examples in its test specifications and curriculum framework. These essential, supporting documents also provide the opportunity for the state to attend to one remaining concern.

Too many indicators still rely on verbs such as "recognize," or "know," which could lead to a preponderance of items on New Jersey's tests that ask students simply to recall information, rather than apply principles, concepts and skills, thereby undermining New Jersey's clearly stated aim to promote critical thinking. Explaining by example what "recognize" and "know" really entail will preserve the level of rigor New Jersey intends for its standards.

ALIGNMENT OF TESTS TO STANDARDS

Having high-quality standards in place is critical to education reform. Having aligned and challenging tests that measure student achievement is equally important.

To determine the degree of alignment between New Jersey's Elementary School (ESPA) and Grade Eight (GEPA) Performance Assessments and its standards, Achieve reviewers systematically analyzed Form A of the 2000 tests, using a protocol based on defined criteria. The following summary identifies strengths of New Jersey's grade 4 and grade 8 assessments and areas for improvement:

- ✓ **Although New Jersey's science tests and standards generally are aligned, test items should more closely assess the performances described in the standards, and the standards should be more evenly assessed.**

New Jersey's standards and progress indicators describe what students should know and be able to do. Overall, there was good agreement between test items and the science standards. However, reviewers found the content match between the test items and the standards to be stronger than the match between the performances demanded by the test items and what the standards called for. Performances are less aligned principally because the state uses verbs in its indicators that stress the importance of hands-on science, such as "investigate," but that do not lend themselves to being assessed by large-scale, paper-and-pencil, predominantly multiple-choice tests. Performance alignment could be improved by framing the indicator so the emphasis on process is preserved, even while the indicator is stated in measurable terms. For example, progress indicator 5.8.3, "Investigate matter by observing materials under magnification," could be restated to focus on what students should learn as a result of the investigation. Another approach is to generate items that directly assess the student's grasp of understanding science process skills. For example, progress indicator 5.2.9 states that students will "communicate experimental findings using words, graphs, pictures and diagrams." This skill is most easily assessed with an open-ended item. However, it could be assessed with an item that supplied data and then asked,

"Which of the following … communicates best how the experiment turned out?" Either strategy would maintain New Jersey's rightful concern that students are able "to do science," and not just recall answers to content questions.

In the long run, reviewers recommend that New Jersey revisit its original intent to develop aligned performance assessment tasks and scoring guides to be administered locally. This is the final building block in an otherwise comprehensive system of science instruction and it goes to the heart of the scientific enterprise. If put into place, it would deepen students' ability to think scientifically and greatly assist New Jersey in realizing its goal of scientific literacy for all of its students.

Reviewers also recommended that in revising its assessments, New Jersey should ensure that the standards are more evenly assessed. Assessments that are limited to 38 items (ESPA) or 48 items (GEPA) and must cover 12 standards with 49 progress indicators at grade 4 and 59 progress indicators at grade 8, must use every item judiciously to end up with a balanced assessment. Not every indicator must be assessed to achieve balance. What is key is that items selected measure the major concepts and performances expressed in the standards. As it stands, some standards and some indicators within standards are overassessed, leaving others — some of which are of greater significance — unassessed. For example, Standard 5.2, which is mainly concerned with science inquiry, is not well assessed.

- ✓ **While New Jersey's assessments have significant strengths, the level of rigor of the ESPA should be raised.**

New Jersey deserves credit for moving away from emphasizing simple recall in favor of asking students to apply knowledge across standards and even, on occasion, figure out how to analyze a relatively complex situation. Another positive feature of the tests is the use of open-ended items, which call upon students to approach a task or problem in its entirety and reason out a response rather than select an answer from alternatives. Moreover, the state set itself a difficult, but worthwhile, task in attempting to have every item assess two standards. It is a way to help bridge the gap between content and process skills.

Reviewers recommended ways in which these features of the tests could be enhanced. Open-ended items have the greatest potential to probe students' understanding, but for them to be used effectively, they should not assess knowledge and skills readily assessed with multiple-choice items. In addition, the scoring rubrics or guides and accompanying examples of scored student work must be well thought out. That was not always the case, especially on ESPA. Reviewers also recommended that the test developers review individual items carefully to determine the degree to which the goal of assessing two standards is realized. In some cases, reviewers were not able to confirm that an item, in fact, assessed more than one standard. When they did find that an item assessed two standards, typically one standard was the primary target. To ensure the public is clear about student proficiency, New Jersey may only want to include primary "hits" in its score reports.

Reviewers noted that, overall, GEPA was a stronger test and contained models of items that could be used to improve ESPA. ESPA contained an unacceptably high number of flawed items, where the distracters in multiple-choice items were so unconvincing it was difficult to imagine a student choosing an incorrect answer. In other words, the correct answer was given away. When items have this flaw, it is not safe to assume that a student who chooses the right answer has a good grasp of the content and skills being assessed. The net result is that the level of rigor was seriously reduced. That said, such flaws are readily correctable by undertaking a careful review of individual items.

CONCLUSION

In sum, New Jersey is developing a sound science education program based on strong standards, assessments, and multiple supporting documents — test specifications, a supporting curriculum framework and sample tests. This work puts New Jersey ahead of states where little attention has been given to science, or where standards exist without aligned tests or supporting curriculum. This is particularly good news given the provisions of the federal No Child Left Behind Act that require states to have science standards and tests in place in the coming years. More importantly, it bodes well for the future achievement of New Jersey's students.

Board of Directors, Achieve, Inc.

Achieve, Inc.
www.achieve.org

400 North Capitol Street, NW
Suite 351
Washington, DC 20001
Phone: (202) 624-1460
Fax: (202) 624-1468