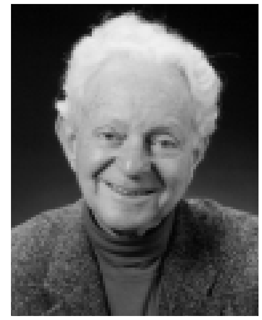
Science Education in the Twenty-first Century: Pushing the Envelope on Student Assessment

April 11, 2001
Aurora, IL



Dear Colleagues:

The Illinois Mathematics and Science Academy, through its Great Minds Program, began Science Education in the 21st Century as a series of dialogues on critical issues in science education. Initiated three years ago to probe issues confronting science education in Illinois, this series began with the dialogue *Physics First? Redesigning the Curriculum for High Schools*, which questioned the traditional structure of the high school science curriculum—biology, chemistry, and then physics. The first dialogue showcased efforts to rethink the curriculum, to reorganize it based on what we know about how students learn, and to change it based on what has emerged as new science content in the past century, since that original structure was put in place. In 2000, our dialogue probed the question: *What will it Take to Deliver a Standards-Based Curriculum?* as we, Illinois educators, began grappling with the expectations for students, driven by relatively new state and national standards in science.

This dialogue, *Pushing the Envelope in Student Assessment*, confronts the challenges of assessing student learning in science and explores the purposes, the tools, and the technology for assessment. It is a conversation about assessment in science and what it should "look like" as we prepare students to become full participants in shaping the story of the new millennium, especially as it relates to acquiring, generating, and using knowledge and understanding for the World.

Under the sponsorship of IMSA's Great Minds Program, our goal is to bring national and international leaders in science, mathematics, the arts, and humanities to IMSA to engage in conversations with our students and our staff, as well as other students and teachers throughout the state of Illinois. Our intention is to turn what we are calling "conversations that matter" into "actions that make a difference."

Common to all of the Great Minds Program dialogues is the assemblage of a distinguished panel of educators, scholars, and leaders who have accepted our invitation to come and think out loud with each other and with us, in real time, and to engage and interact with participants in exploring the critical issues of learning and teaching.

We hope that the issues raised in the following pages will inspire you to work together with us to "push the envelope on student assessment" so that we may integrate the very best ideas of teaching and learning into the classroom and move education forward to benefit all of our children.

Sincerely,

Stephanie Marshall

Stephanie Pace Marshall Ph.D.
President
Illinois Mathematics and Science Academy

Leon M. Lederman

Leon M. Lederman, Ph.D., Nobel Laureate
Founder and Resident Scholar
Great Minds Program
Illinois Mathematics and Science Academy

Bruce Alberts

Bruce Alberts is president of the National Academy of Sciences and chair of the National Research Council, the principal operating arm of the National Academy of Sciences and Engineering. Alberts obtained his Ph.D. from Harvard University in 1965 and is a respected biochemist recognized for his extensive molecular analyses of the protein complexes that allow chromosomes to be replicated. He is one of the principal authors of *The Molecular Biology of the Cell*, now in its fourth edition, considered the leading advanced textbook in this field and used widely in U.S. colleges and universities. Alberts is committed to the improvement of science education and helped to create City Science, a program for improving science teaching in San Francisco elementary schools.

Lynne Haeffele Curry

Lynne Haeffele Curry is chief deputy superintendent of the Illinois State Board of Education. She received her graduate degree from Illinois State University and taught science at Bloomington High School for thirteen years, the last five of which she also served as science department chairperson. In 1985, Curry represented Illinois in the Teacher in Space competition. She joined the State Board of Education in 1990, holding the post of supervisor of the Center on Scientific Literacy, establishing the Center for Educational Technology, becoming division administrator for Strategic Planning and Budget Management, and was project manager for the Illinois Academic Standards Project that resulted in the adoption of the Illinois Learning Standards in 1997. In 1997 Curry was selected Executive Assistant to the State Superintendent and subsequently served as Deputy Superintendent for Educational Programs and for Standards, Assessments, and Accountability.

JoEllen Roseman

JoEllen Roseman was curriculum director for Project 2061 of the American Association for the Advancement of Science, at the time of the dialogue. In this capacity she was involved in the design, testing, and dissemination of Project 2061's science literacy reform tools. She was also in charge of overseeing Project 2061's middle and high school science textbook evaluations and developing the new reform tool, *Resources for Science Literacy: Curriculum Materials Evaluation*. In 2001, she assumed the position of Acting Director. Before joining Project 2061, she was a member of the faculty in Arts and Science and Education at Johns Hopkins University where she designed and developed graduate programs for secondary school science teachers and for prospective teachers. Roseman has also served on the Board of Directors of the Biological Sciences Curriculum Study (BSCS).

Richard J. Shavelson

Richard J. Shavelson is professor of education and psychology (by courtesy) at Stanford University. He served as the I. James Quillen Dean of the Stanford University School of Education from 1995-2000. Before joining Stanford, he was dean of the Graduate School of Education at the University of California at Santa Barbara (UCSB). Shavelson is also a former president of the American Educational Research Association. His research focuses on linking assessment methods with a working definition of achievement that includes declarative, procedural, strategic, and meta-cognitive knowledge. Co-author with Professor Noreen Webb of the book, *Generalizability Theory: A Primer*, Shavelson's other psychometric publications include research on the dependability of performance assessments used in work and education.

Introduction

Not everything that counts can be counted.
Not everything that can be counted, counts.
Albert Einstein

Limitations of current assessments and how results are used impact what is measured and reported regarding student learning in science. Large-scale assessments are costly and constrained by needs for efficiency and fairness. Lacking measures beyond those focusing on science vocabulary and concept recognition, most assessments in science inadequately sample the depth and breadth of science as expressed in state and national learning standards. The standards call for greater attention to physical sciences, earth and space sciences, and inclusion of scientific inquiry, issues related to science, technology and society, and technological design in the science curriculum. Commensurate actions are lacking to help unburden already crowded curricula and to develop assessment tools and measures in these areas for classroom and large-scale assessments. The result is to perpetuate an emphasis on what can be tested through existing narrow assessments. Using large-scale assessment as the primary means of accountability stresses reporting and sanctions rewards rather than promoting actions needed to improve student learning through redesigned curricula, professional development, teacher recruitment, and development of systems that support an increase in well qualified science teachers. Finally, there is great need for more educators who understand standards-based education reform and who can use the tools, research, and experiences of science education reform to improve student learning.

On April 11, 2001, IMSA convened the Great Minds Program Dialogue: *Science Education in the Twenty-first Century: Pushing the Envelope on Student Assessment*. Panelists were asked to consider several key questions:

- What do you see as the most promising developments in assessment? Why?
- What would assessment look like that captures the depth and breadth of the state or national science standards?
- How can we reconcile the information needs of various audiences with the capacity of the present education system?
- How can assessment become more intrinsic to teaching and learning?
- What external conditions (legislative, judicial, etc.) promote or inhibit good assessment practices? How?

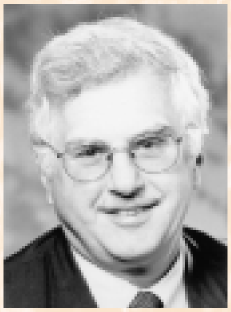
The members of the distinguished panel were:

Bruce Alberts, President of the National Academy of Sciences, Washington, D.C.
Lynne Haeffele Curry, Chief Deputy Superintendent of the Illinois Board of Education, Springfield, Illinois
JoEllen Roseman, Acting Director, Project 2061, American Association for the Advancement of Science, Washington, D.C.
Richard J. Shavelson, Professor of Education and Psychology (by courtesy), Stanford University, Stanford, California
The panel dialogue was moderated by **Leon Lederman**, Resident Scholar of the Great Minds Program, and **Stephanie Pace Marshall**, IMSA President.
Michael Lach, Special Assistant to the Chief Technology Officer, Chicago Public Schools, Chicago, Illinois, moderated questions at the closing session of the dialogue and provided reflection.

This paper attempts to capture the essence of the day, to synthesize the ideas presented by panelists and practitioners, and to provide Internet resources and reference information. The planning team for this dialogue included Dr. Raymond J. Dagenais, Dr. Michael J. Palmisano, Dr. Steven Rogg, and Dr. Judith A. Scheppler. IMSA thanks Drs. Scheppler and Rogg for their writing efforts in synthesis of this Great Minds paper. We invite you to contact us at greatminds@imsa.edu or www.imsa.edu to be placed on our mailing list, make observations, ask questions, or propose topics for future dialogues.

Bruce Alberts

President, National Academy of Sciences, Washington, D.C.



Currently, there is a focus on accountability in our schools and there will be high-stakes reading and mathematics assessments. Science will also need to be assessed, but what kinds of science assessments are needed? As stated in the National Science Education Standards, we need to place "less emphasis on assessing what is easily measured," and "more emphasis on assessing what is most highly valued." In other words, we need to put less emphasis on assessing scientific knowledge and more emphasis on assessing scientific understanding and reasoning. Most of our present science assessments need to change. *Alberts spoke on the topic of science assessment, based on his own experience as a university and medical school professor. He related the following anecdotes and examples of science assessments to underscore his message that the assessment system, especially for science, needs to change.*

Anecdote I: Assessment in a Biochemistry Course

An illustration of the effect of assessment on student learning occurred while teaching biochemistry to medical students. A change was made from an easy scantron-based multiple choice test to one that consisted of half of the questions being short essays. The year that that occurred, there was a dramatic change in the kind of questions the students were asking during and after class. Their whole attitude towards the course changed. It had such a striking effect, that it made one realize how critical exams are to student learning, especially exams that count. If these students were taught just to memorize facts, then its possible that the best students had not been selected to become the best doctors.

Anecdote II: The Princeton Review for the SATII Exam

The SAT II test in biology is a major national exam widely used today; many high school teachers will need to prepare their students for this exam. A quote from the Princeton Review for the SAT II exam reads:

"We'll show you that you really don't have to understand anything. You just have to make a couple of simple associations, like these … Aerobic respiration with: presence of oxygen more ATP produced … Anaerobic respiration with: absence of oxygen less ATP produced … When we get through, you may not really understand much about the difference between aerobic and anaerobic respiration. But you don't have to, and we'll prove it … Whether or not you understand your answers, the scoring machines at the Educational Testing Service will think you did. Their scoring machines don't look for brilliant scientists and they don't look for understanding. Stick with us, and you will make the scoring machines very happy."

It seems funny, but in fact it is tragic because we are creating a whole generation of students who have an attitude that education is trivial and irrelevant. And this is correct, if this is what they are talking about as education. Since adults produce this exam, we have a major problem.

Example: The Maryland School Assessment Program

There may be some promising examples on which to build. For example, Maryland has developed the Maryland School Assessment Program, which involves a week of testing every year for third, fifth, and eighth graders. Rather than only compartmentalizing science, mathematics, reading, and writing, they test for multiple abilities at once. As an example, the following question is one that was asked of all Maryland third graders:

The problem: Your teacher has received a bouquet of flowers and is having trouble with them. The leaves are drooping, and the flowers look sick. You decide to do an investigation to discover what might be wrong with them. Students must then perform the following tasks:

(1) Read two articles about plants and their stem system.
(2) Write an essay explaining how you would study your teacher's flower to determine what's wrong with it.
(3) Draw an illustration that would help other students understand your investigation.
(4) With a partner, use a magnifying glass, look at the cut edge of a bottom of a celery stalk (which is used in place of the flower), make a list of things you observe about the stalk, break the stalk, and describe what you see.
(5) Draw and color a picture of what you think will happen to this celery if it sits in red dye overnight. Explain why you think so.
(6) On the next day, study the celery that was soaked overnight in the red dye. Write a paragraph to explain how the celery is the same or different from what you predicted yesterday.
(7) Write an essay explaining why a scientist might want to do more than one investigation when trying to answer a question about science.
(8) And last, write a note to your teacher telling what you have learned about flowers and how to take care of them.

The Maryland assessments are written by teachers and graded by teachers over the summer, and they represent an important professional development exercise for the teachers. This type of exam would seem to test for the type of abilities that we want kids to acquire to prepare them for the real world. And it would seem to make school meaningful to them. With these kinds of questions, parents should also be able to appreciate the relevancy of school to their children's lives, and see its importance for getting a skilled job.

- *Assessment Theorem:* **What is measured in high stakes assessment has a profound effect on human behavior.** *Corollary:* **We must be exceedingly careful to make sure we measure what counts.**

Research Needs to Support Change

The kind of assessment used in Maryland stands in stark contrast to testing for the memorization of the parts of the cell, or any of the other types of science knowledge that are more commonly assessed in high-stakes, large-scale assessments. A virtue of our decentralized education system is that many large-scale experiments are being carried out in our nation. One would urgently like to know how the Maryland assessments affect what teachers teach and what students learn—as compared to similar high-stakes examinations of a more standard kind that are given in some other states. Unfortunately, we don't know.

In general, we have failed to take advantage of our uniquely decentralized education system. We have an opportunity here, if we can bring the best science to bear in a serious evaluation of the many interesting real-world experiments being carried out by various states. With regard to assessments, we clearly would benefit from good scientific evidence regarding what kinds of tests, that are being given today, are promoting the type of science education outlined in the *National Science Education Standards*, if any.

One of the things we need to do, if we are to get serious about education, is to conduct the high quality research that it takes to build a continually improving system, *based on evidence*. A well-known fact is that if you give any kind of multiple use exam, which is a high-stakes exam, teachers learn how to teach to that exam. The test scores will go up in about ten years, so you can claim victory. But it doesn't mean victory. It doesn't mean anything at all. All it means is that the teachers have become more sophisticated at teaching to that exam because their job is at stake if their students don't achieve good scores. It also may mean that the kids know a lot less about science then they knew before, because this exam is probably encouraging the wrong kind of teaching.

The National Academy of Sciences

The National Academy of Sciences (NAS) has a web site (www.nationalacademies.org), which contains 2,000 full texts, online. They can be accessed freely and many books can be ordered on the web. By clicking on "education" on the bottom left hand side, you get links to many major reports, including the *National Science Education Standards*. Recently, the NAS has been taking a closer look at the national science standards:

- **When we measure some things and not others that which is not measured tends to get neglected.**

- The National Academy of Sciences has produced an excellent, high quality guide, *Inquiry and the National Science Education Standards*, for teachers that explains what we mean by science inquiry, with many kinds of examples of inquiry. This material points out that inquiry need not be hands-on, but can be conducted both without hands-on activities and with hands-on activities.
- Similar kinds of reference materials need to be developed for college faculty, because one of the major focuses of the NAS is to change the first year of science courses in colleges and universities. A change of science curriculum and assessment at the university level is also critical to improving scientific literacy, overall; overhauling primary and secondary education is not sufficient.
- The NAS produced a report, *How People Learn*, which takes into account the last thirty years of the psychology of learning and translates this into what that means for our schools. This is now being used as a textbook, available in paperback form, for many pre-service teacher education programs.
- The NAS has conducted studies on assessment and are especially concerned with high-stakes testing for tracking promotion and graduation. On the NAS web site are several specific reports on assessment; of note is *Knowing what students know: The science and design of educational assessment*. Its main point is that some of the things that we are trying today, testing for what we really want students to know, can be done if you combine what we know about how people learn, (cognition) with the new technology (computer technology for doing testing). There are some real examples of assessments that show us that we can do different testing than what we have been doing, and make it scaleable. It is also clear that we do need more research and development in this area.

We need widespread recognition that any high-stakes examination that is given over multiple years will change the teaching of science in a way that leads to successively higher test scores: as time goes by, teachers will become more expert at teaching to the test. What we urgently need to know is whether these improved test scores are reflecting an improvement in science learning by an average student. To obtain this information, we will need to devote new resources and new scientifically trained researchers to a much more serious and focused set of research efforts than those commonly available today.

JoEllen Roseman

Acting Director, American Association for the Advancement of Science,
Washington, D.C.

A Phi Delta Kappa/Gallup Poll (October 24, 2000 *Washington Post*) showed that despite recent reports that parents are happy with their own child's public school, they are criticizing education in general. *Roseman contends that even the best students are not learning important ideas and skills in science. Her expertise and insight about science assessment stems from directing the American Association for the Advancement of Science (AAAS) Project 2061 textbook and curriculum evaluation. She shared some evidence which supports the lack of proper science education, but then went on to illustrate that Project 2061's textbook evaluation shows that middle school and high school biology texts are incoherent and offer little support for instruction.*

Some Evidences

In a prestigious Maryland county, even high achieving eighth grade students were unable to explain everyday phenomena using ideas from the kinetic molecular theory. For example, hardly any of the "A" students were able to describe boiling or dissolving in terms of molecules or to explain a house burning to the ground in terms of the conservation of atoms.

- **Even the best students are not learning important ideas and skills in science.**

In a similar study, most top biology students in another Maryland county revealed serious misconceptions about natural selection—misconceptions that would make them unable to think productively about issues like the development of antibiotic resistance in bacteria (which appeared in the news the same week Project 2061 released findings of the textbooks evaluation).

And even graduates from MIT and Harvard University were unable to explain where the mass of a tree comes from, such as a maple tree, which starts out as a tiny seed and grows into a tree weighing hundreds of pounds. (It comes mainly from the carbon dioxide in the air.)

Project 2061 Textbook Evaluation

Project 2061 undertook a biology textbook evaluation. The textbooks were analyzed for their treatment of four topics central to biology: cell structure and function, matter and energy transformations, molecular basis of heredity, and natural selection and evolution.

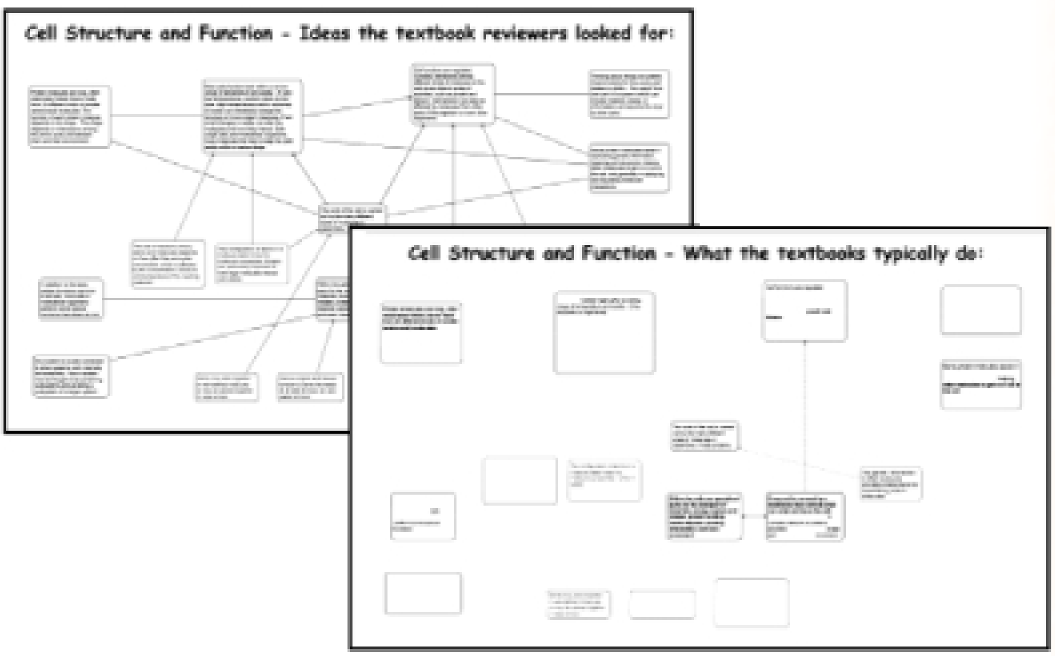
Reviewers had to answer two main questions for each textbook:

1. For the **Content Analysis**, Project 2061 asked: Are the key ideas within each topic treated and are they coherently related to one another?
2. For the **Instructional Analysis**, Project 2061 asked: Does the textbook help students learn the key ideas and does the teacher's guide help teachers teach them?

For students to learn the key ideas, the answers to both of these questions must be "Yes." For each topic, a set of related key ideas were selected from *Benchmarks for Science Literacy* and the *National Science Education Standards*, on which most states claim to have based their own standards.

The main ideas of a topic served as a basis for analysis and were mapped to prerequisite knowledge and closely related topics that would strengthen student understanding. If a textbook addressed all the ideas and made clear their connections, it would likely be telling a coherent story of the topic.

The reviewers looked for these ideas and connections in each teacher's edition. They examined everything relevant to the topic in the learning objectives stated in the text, such as student readings, discussion suggestions for the teacher, lab investigations, questions, and sample answers. Whenever they found something "on target," they underlined text in the box. When they were done, any text not underlined was "erased" for that book, indicating that the idea or part of the idea was not treated at all. If connections were found between two ideas, the arrow was highlighted. If a connection was not found, the arrow was "erased" for that book. The figure illustrates, schematically, results for cell structure and function.



Maps such as these were used to identify specific content and relationships that reviewers expected to see in textbooks (background) and contrast with what the reviewers actually found (foreground). (Source: Project 2061)

Content Analysis Findings

Do biology textbooks treat the key ideas and connect them to one another? A map of what the reviewers actually found in the textbooks with regard to the cell story, shows the following:

- **only four of the boxes are full of text**—which means that only a few of the complete ideas are found in textbooks. For example, materials do present the idea that cells have specialized parts for transporting materials, for building proteins, and so on. Textbooks all have a diagram of a cell that names the parts and is followed by several pages of text describing the functions of those parts.
- **five boxes are empty**—because the books don't address the ideas. For example, none of the books treat related ideas about systems from a point of view of systems design. Yet doing so, could help students think of the cell as having parts functioning together as a whole, rather than as just a bag of miscellaneous parts and functions.

- **a few boxes show some fragments**—because only parts of ideas are addressed. For example, a few textbooks indicate that "some protein molecules help other molecules to get in or out of the cell" but don't indicate that proteins, in fact, do nearly **all** the work of the cell-replicating genetic information, building and repairing cell structures, releasing energy, and generally regulating molecular interactions.
- **there are hardly any arrows**—even when books do treat two related ideas they don't relate them to one another. For example, even though a book might mention that proteins do the work of the cell, and claim that DNA provides instructions for assembling protein molecules, the book doesn't relate these ideas to one another—by noting, for example, that cells function according to instructions coded in DNA.

- **Despite recent reports that parents are happy with their own child's public school, they are criticizing education in general.**

Even though textbooks are not treating many of the key ideas or connections among them, they still have one to two whole chapters on cells. But the information is presented as a laundry list of facts. So students who take biology (which is nearly everyone) will likely come to view the cell as just a bag of parts with strange sounding names rather than a living thing.

It's important to note that reviewers were very lenient about giving credit for a content match. All a textbook needs to do to get credit for a content match on any idea is to mention the idea or connection in a sentence. But this is woefully insufficient to help students learn these ideas, some of which are quite abstract. To look at whether students could actually learn the ideas presented, the team carried out an instructional analysis for each key idea, regardless of whether it was treated extensively or just mentioned.

Instructional Analysis Findings

Do the biology textbooks help the students learn the key ideas and does the teacher's guide help the teacher to teach them well? The broad categories of the criteria used to analyze the quality of instruction provided for the content are listed in the table. It is important to point out that the analysis doesn't simply look to see whether textbooks use some good instructional strategies but whether the strategies are actually focused clearly on the key ideas. All (or nearly all) of these instructional criteria must be met for students to learn well. Just as all the pillars are needed to support a structure, if any of the criteria are missing, the whole structure may collapse.

The instructional analysis findings for the topic natural selection are summarized for ten textbooks. (The Project 2061 website has five such charts-one for each of the four topics examined and the fifth one showing the average across all four topics).

Category I: Providing a sense of purpose. Three books do some promising things to provide students with a sense of purpose.

Category II: Taking account of student ideas. The textbook authors do not take advantage of the extensive research base on common misconceptions students have about natural selection (and other key ideas) and appear to have been unaware of it when designing activities.

Category III: Engaging students with relevant phenomena. Half of the textbooks include an acceptable variety of phenomena that can be explained by the key ideas—this means they show examples of diverse organisms having similar skeletons, patterns of early development, and DNA sequence that can be explained by them having descended from a common ancestor. And three of these books provide descriptions and diagrams to give students a vicarious sense of the phenomena. These phenomena could, if used well, help make the key ideas of natural selection credible to students.

INSTRUCTIONAL ANALYSIS CRITERIA
I. PROVIDING A SENSE OF PURPOSE
Conveying unit purpose
Conveying lesson purpose
Justifying lesson sequence
II. TAKING ACCOUNT OF STUDENT IDEAS
Attending to prerequisite knowledge and skills
Alerting teacher to commonly held student ideas
Assisting teacher in identifying own students' ideas
Addressing commonly held ideas
III. ENGAGING STUDENTS WITH RELEVANT PHENOMENA
Providing variety of phenomena
Providing vivid experiences
IV. DEVELOPING AND USING SCIENTIFIC IDEAS
Introducing terms meaningfully
Representing ideas effectively
Demonstrating use of knowledge
Providing practice
V. PROMOTING STUDENT THINKING ABOUT PHENOMENA, EXPERIENCES, AND KNOWLEDGE
Encouraging students to explain their ideas
Guiding student interpretation and reasoning
Encouraging students to reflect on their own learning
VI. ASSESSING PROGRESS
Aligning assessment to goals
Testing for understanding
Using assessment to inform instruction

Categories IV: Developing and using scientific ideas, and V: Promoting student thinking about phenomena, experiences, and knowledge. However, none of the books provide much help in promoting student thinking about the phenomena or helping them develop and use the scientific ideas to explain the phenomena. While one book does a decent job of encouraging students to explain their *own* ideas that might be relevant, none of the books provides adequate guidance in helping students interpret the phenomena in terms of the *scientific ideas*. Phenomena are not developed to help students see the ideas as credible or to appreciate how they can help to explain everyday events reported in the news, such as the story about our losing the war on germs that recently appeared in the *Washington Post*. So most of the phenomena mentioned in the books are likely to have little impact on learning.

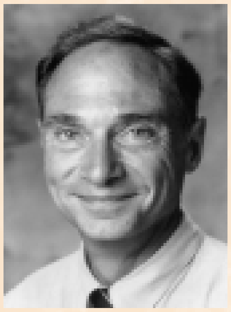
Textbooks are not developing an argument for students about any of the ideas. They are not stating their assumptions and are leaving out so many steps in the argument that even highly motivated students are unlikely to be able to make connections. And the consequences of this may be even more insidious than appears. Students who do poorly are obviously short-changed. But so are the students who *appear* to succeed.

> *The textbooks, by their sheer bulk and by their fancy displays, convey the impression that they cover the subject. They imply that, if you learn the book and pass the test, you know the material. So students who get good grades think they have learned something but some, in fact, really haven't.*

Project 2061's findings reveal that textbooks rarely attempt to present a coherent and comprehensible story. Even when they do, they don't support students in learning that story or teachers in teaching it. As a result, the assessment enterprise faces both substantive and political challenges. Depending on how it is implemented, testing can either perpetuate the problem, make it worse, or contribute to solving it. A first and necessary step is to ensure that the enacted curriculum is firmly aligned with the learning standards. Only then is meaningful assessment of student learning possible.

Richard J. Shavelson

Professor of Education, Stanford University,
Stanford, California



Assessment in science needs to reflect students' understanding. We all seem to be pretty much in agreement on this. But first we need to decide what it means to *"achieve in science."* Until we decide what this looks like, we will still have the type of substandard texts, which Dr. Roseman described. And these texts are in perfect agreement with current large scale, high-stakes standardized testing, described by Dr. Alberts, which ask kids about pieces of information or individual facts. *Shavelson built on the remarks of the previous two speakers, espousing the need to teach for conceptual change, describing what he feels achievement in science looks like, and how various types of assessments may be used to test students' knowledge.*

There is a disjuncture in testing, between the externally mandated summative assessments and the formative assessments that teachers need to use in order to improve their teaching practices. It is clear that President Bush's educational agenda will include testing as the responsibility of each state. It is also clear, however, from his tenure as Governor of Texas, that the types of tests he supports will not reflect what most teachers believe is important, teaching for understanding. Teachers need to understand the ways in which concepts in science have evolved in their students' minds, and then to bring their students' understandings of science into a closer fit with what scientists currently understand about the way our natural world works.

- **Assessment in science needs to reflect students' understanding.**

There needs to be an alignment between summative and formative assessments. If this does not occur, then we're always going to continue in our present mode of ineffectual teaching and learning. And, in the end, these summative kinds of assessments are going to drive the entire system and we are not going to see radical changes in testing because we have a technology that is very efficient and effective for this. This means that we're also not going to see big changes in textbooks.

A Conceptual Framework for Scientific Knowledge

Defining the knowledge about science that we want students to know is critical to being able to develop effect assessments. In some ways, it is the harder task since creative educators have already developed many types of assessments. Using cognitive theory, different types of knowledge that experts possess have been identified and these might serve as goals for students in their learning and doing science. This conceptual framework for scientific knowledge divides this knowledge into declarative, procedural, schematic, and strategic knowledge. This knowledge framework can be linked to cognitive and "situated" theories of learning and cognition. Moreover, these different types of knowledge have been linked to different types of assessment techniques.

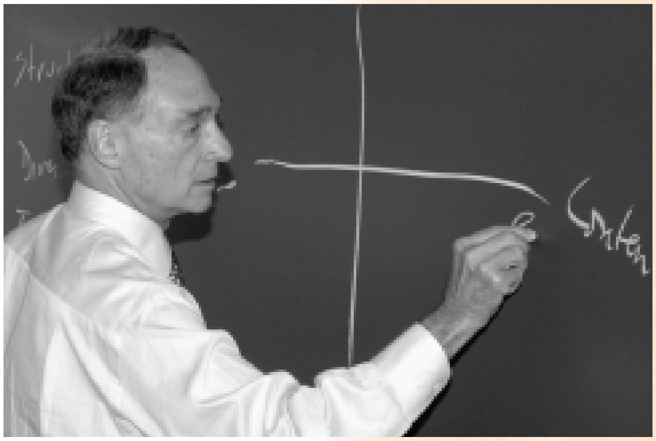
Declarative knowledge is domain-specific content knowledge and consists of facts, concepts, and principles; having to do with "learning that…" Declarative knowledge is important because *learning that* a molecular explanation can explain a variety of disparate things is important. *Learning that* the change of seasons isn't random, but systematic and that there are ways of explaining it, are important. These kinds of "learning that," basic kinds of understandings, are very important. For declarative knowledge to be "usable," the bits of information need to be interrelated conceptually. Experts' declarative knowledge, for example, is highly structured. "Learning science" has been described, at least in part, as a process of building an increasingly sophisticated knowledge structure; that is, as a process of becoming expert in a science domain. Current paper-and-pencil achievement tests do a poor job of measuring the structural aspect of declarative knowledge.

A second kind of knowing that is an important assignment in science is learning how to carry out a well-controlled investigation, called **procedural** knowledge. This includes students' collecting data, representing the data in some format, and interpreting what they see. After all, scientists conduct investigations, testing curiosities, hunches, and theories. They test their ideas, for example, by manipulating some variables and controlling others to gather empirical support for their hunches, and theories. Shouldn't our conception of science achievement include the knowledge and skills needed to conduct such investigations? We all certainly want our students to know, understand, and use procedural knowledge. Again, current paper-and-pencil exams are inadequate assessments of procedural knowledge. Students must conduct performance assessments in the form of hands-on investigations, perhaps carried out with computer simulations or in mini, portable laboratories in order to assess procedural knowledge.

- **…we need to decide what it means to "achieve in science."**

Additionally, there is **strategic** knowledge, knowing when, where, and how to apply what you know. Often kids are seen formula stuffing; they have three variables and they need three numbers so they stuff the numbers in from the problem statement and they go with it. You have to understand when the knowledge that you have applies to a particular situation. Experts combine concepts and procedures in the form of rules for action under certain task demands and work conditions, and they are very good at that. They take a look at three different problems and say that underlining this is Newton's second law, and it brings three items together and therefore the fundamental law applies to all three problems. The result is a set of alternative plans to solve a problem.

Finally, we talk about **schematic** knowledge, an understanding or mental model of how systems work. For example, your mental model explains the change of seasons. It is related, certainly, to your declarative knowledge. Experts seem to structure knowledge into the form of mental models, and then they are able to use these models of underlying principles and relationships to bring their declarative and procedural knowledge to bear on solving a new problem or testing a new hypothesis. Some progress has been made through interviews, sketches, and even multiple-choice tests that identify students' mental models.



Teaching for Conceptual Change

Within this conceptual framework, however, the most important and toughest task in teaching science is teaching for conceptual change; taking students' mental models of how they view the physical world around them and then bringing those models more in line with the way scientists think. For example, everyone "knows" that if you don't keep pushing an object with force it will stop. And therefore, naturally, to keep things moving, you have to have a continuous force. That is what Aristotle thought. You wouldn't have to change the question of moving objects, but you need to change students' understanding of it. Teaching for that conceptual change is very hard, because you are asking kids to change their beliefs, not just their knowledge, but what they believe in, what's functional in the real world for them. Assessing the kind of schematic knowledge that kids have helps determine the conceptual changes that need to take place.

We can talk about the characteristics of scientific knowledge. How much of the knowledge do you have? How well structured is that knowledge? If you are thinking about declarative knowledge, often times we test for how much knowledge students have in bits and pieces. But we can also ask: how well structured or organized is that knowledge? We know that experts have very well structured knowledge. They see the forests and the trees and how those things are all related to one another; novices don't do that. Taken all together, the different types of knowledge, the amount of knowledge, its organization, and how it is used defines what it means to achieve in science. This view is consistent with other experts, including Roger Bybee, who have worked on the national science standards. It is also consistent with what cognitive psychologists believe as well as many people working in the assessment arena.

- **There is a disjuncture in testing, between the externally mandated summative assessments and the formative assessments that teachers need to use in order to improve their teaching practices.**

"Situated" or "socio-cultural" learning theory builds on many of the ideas of cognitive theory, but also recognizes the social nature of learning. That is, learning frequently occurs among people within a particular community (e.g., science classroom). And that learning consists of acquiring the "tools" and language of the community, such that competence is displayed by a student's capacity to interact and communicate with peers in conducting an investigation or in solving a problem. Learning progresses as a student increases her capacity to do more advanced things or recognizes the generality of a principle learned in one setting to other settings. This situated theory recognizes the importance of teachers' formative assessment in classrooms-their informal observations, interviews, and discussions contributes to the community and also provide valuable information about how individuals and groups of students are progressing.

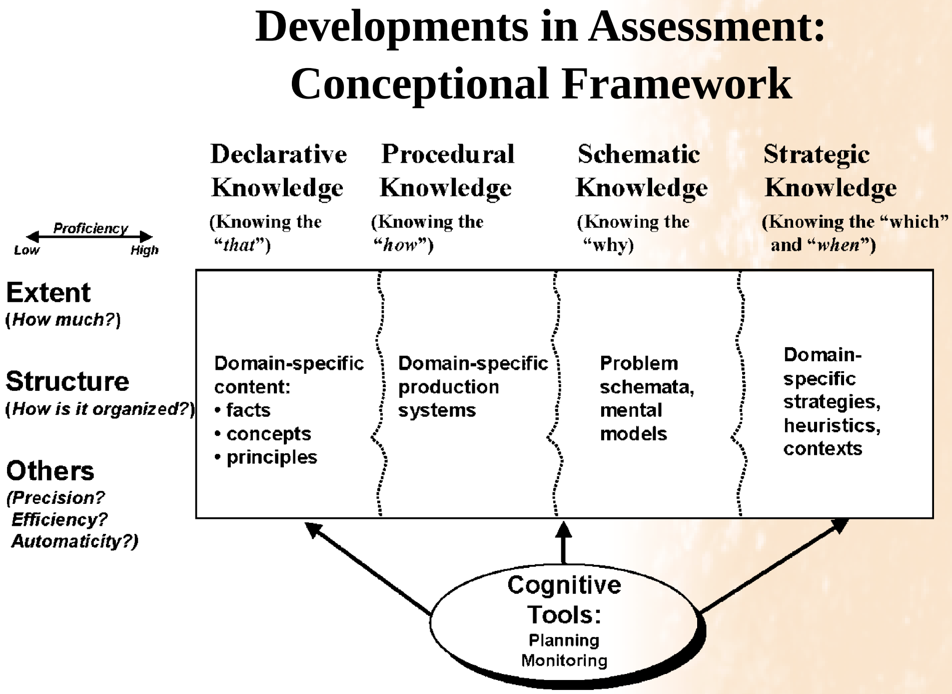
Assessments and Conceptual Framework

Finally, getting back to the dialogue topic of assessment, one asks the question: Can we link particular types of assessments to this conceptual structure? Shavelson is currently engaged in empirical research, which evaluates the capacity to make these links between assessments methods and the kinds of understandings that we want students to possess. Some examples of these assessment methods are:

- **A multiple choice test.** These do an excellent job of measuring the amount of declarative knowledge a student has. If this is all you focus on, then you get textbooks like Dr. Roseman and Project 2061 have evaluated.

- **Concept mapping.** You can begin to look at the structure of students' knowledge and their understanding of how key ideas in the domain are related to each other by using concept maps. There are other ways at getting at the structure of knowledge, but concept maps are straightforward and easy for teachers to use. They are also an extremely effective teaching devices.
- **Performance assessment.** In a performance assessment, you give the students a real world situation; you pose a problem, give them some equipment, and let them conduct an investigation.
- **Developmental assessment.** A developmental assessment helps identify mental models, trying to determine students' understandings of how a phenomena works.

There are two promising developments on the horizon for science education assessment. These are technological and methodological breakthroughs. If we are to realize the promise of alternative assessments, we have to develop a *technology* that makes them cost-efficient for both classroom and large-scale use. In our work on performance assessments, we have identified different types of these assessments that have shortened the assessment-development process greatly. Moreover, we have put a few of the assessments on a computer platform and have studied the link between hands-on and computer-generated performance assessments. *Methodologically*, there are three major contributors to dealing more effectively with the complexity of alternative assessments. One is the application of a sampling theory of measurement, generalizability theory, for locating sources of measurement error and testing validity challenges. The second is qualitative item-response theory. And the third is Baysian inference networks. **In short, extensive research in cognition and conceptual change, coupled with both technological and methodological advancements, suggests that a better future for more meaningful assessment of student learning is now possible.**



Lynne Haeffele Curry

Chief Deputy Superintendent, Illinois Board of Education,
Springfield, Illinois

As a standards-led system began to take a foothold, it served to shine a spotlight on all that was inadequate in the old system. One of the areas to receive the first glare of that spotlight was assessment, at the state, district, and classroom levels. Old practices, like springing "pop quizzes" consisting of randomly selected "trick" questions, grading "on a curve," or using paper-and-pencil tests to measure performance skills, became very obviously out of sync with the standards movement. However, we found that no matter how well teachers and administrators "aligned" their curricula with the Learning Standards, they still lacked training and expertise in embedding good standards-based assessments into daily instruction. *Curry's remarks on science assessment practices stem from her vast science teaching experience and her more recent roles in helping lead the Illinois State Board of Education (ISBE). ISBE has instituted a program of professional development aimed at improving assessment within a standards-led education system. Concurrently, they have been looking at Oregon and South Dakota's on-line state assessment system as models for change in Illinois. Curry began her discussion by relating two stories in illustration of classroom teaching practices.*

- **What we are looking at is changing an educational culture in which almost all of us were raised and educated.**

Curry's own epiphany came when she realized that her job as a teacher was not completed until ALL students had reached some important level of scientific understanding, and not until she had the evidence proving that they had reached it. Previously, assessments had been used in a much different way and for entirely different purposes: sorting, ranking, rating, rewarding, punishing. This, coupled with the widespread belief that not all students can learn, caused teachers to throw out lots of knowledge, hoping it would stick. Also, it would serve to select and find only those students who reached the very pinnacle of the assessment scale.

Anecdote I: The Bell-Shaped Curve

Unfortunately, these beliefs have been well ingrained into our children. Upon placing a bell shaped curve on the board on the first day of school, Curry's students would correctly identify it and even label the x-axis with percent correct grade distribution and the y-axis with number of students. They would then split the distribution into a typical small number of "A's" while most of the "students" portrayed in this graph received a grade of "C." When asked how many of them *wanted* to receive an "A," they all raised their hands. When asked how many of them actually *would* receive an "A," they began to distribute themselves out among the bell shaped curve. Asked how they decided what grade they would receive, they made statements such as: "I never did well in science," or "I don't like science," or "I love science."

Anecdote II: Teaching to the Test

This is a story of assessment; actually, assessment gone wrong and its implications. The Illinois state science test was composed of multiple choice questions, geared to sample whether students were learning the information stated in the very broad Illinois Learning Goals, one of which was *know all the important science stuff*. Teachers were speculating on what exactly that sixty-question test would cover. When the results were released, one very small, rural school district had scored far above all others in science performance. A visit to this tiny school was quickly arranged to discern the amazing science teaching and learning that was occurring. And it was truly amazing, as there was only one science teacher, teaching all subjects and grade levels, in a very run-down school without running water, laboratory equipment, or hands-on experiments and only a few old textbooks, a blackboard, and paper and pencils. But classroom observation, while showing a very hard-working and dedicated teacher, did not show innovative teaching methodologies. Rather, the students were simply memorizing a bunch of *stuff*. What was occurring? It turns out, that the teacher was taking the rejected test questions, widely published each year, and teaching the students the basic psychometrics of test taking: how to psych out the test! The saddest part of this story is that the teacher received a congratulatory letter, copied to his board of education, probably ensuring that effective and much needed teaching materials would never be obtained. Assessments have been misused for many years and we are only now beginning to realize the damage that has been done and how critical it is to change the educational mind set.

Illinois Professional Development Begins Improving Assessments

High-quality professional development for classroom teachers and school administrators is occurring in Illinois. The Illinois program took almost two years to develop, and is now operating through every Regional Office of Education and Intermediate Service Center in the state. Consultation from Rick Stiggins and other classroom assessment experts contributed to an intensive, classroom-embedded, and long-term learning process for school teams. While formal evaluation is not yet completed for the first year, anecdotal data indicates that real "light bulbs" of understanding are turning on among the participating educators.

ISBE believes strongly that when state standards *are* used in professional development design, especially professional training on aligning classroom instruction and assessment, the results will show rapidly accelerating student achievement in relation to those standards. This is in stark contrast to the March 21, 2001, edition of Education Week, in which the National Staff Development Council reported that of its members recently polled, "barely half were in districts where state performance standards were used as a basis for designing staff training."

On-line Assessment: Breaking the Mold

Illinois is also learning from Oregon and South Dakota, who have begun a transition to an on-line state assessment. When the Illinois Academic Standards Project staff spoke of their vision of a standards-aligned state assessment, that vision included testing that, while "standardized" in the sense that all students take the tests, could be part of a customized delivery model by using computer technology.

Imagine this: a fourth grade teacher, working with her twenty-two students, decides that five of them are ready to "test out" on the Illinois Learning Standards related to computation. She escorts them to the computer array in the school's learning resource center. A proctor helps them enter their names, passwords, and their requests to take the test questions related to computation on-line. The system verifies the students ID's, selects an appropriate array of questions from the test bank, the students enter their answers (even showing their work on the screen), push a button, and send in their electronic test forms. Within a few hours, the results are back and of course, the students meet or exceed standards, because the teacher assured their competency before they ever logged on. The results are stored in the students' files, to be aggregated with scores for other standards that the students master throughout the year.

Is this scenario impossible? Not according to at least two states, which have not only passed legislation toward this vision, but also which have successfully piloted electronic test administration. While their first administrations will not be as sophisticated as the scenario above, they are making bold first steps to "break the mold" of pencil-and-paper testing in a mass administration model, as we have had for most of the twentieth century.

Another potential plus of electronic administration is the potential for true "performance" assessment on-line. Most of us are well aware of the amazing power of computer simulations: their realism, the ability to watch components again and again, the ability to interact with the medium, and the ability to "save" process steps and results. This technology can radically alter our thinking about the feasibility of performance assessment within a massive state testing system.

Social and Political Factors Inhibit Good Assessment Practices

The general public has a great concern for accountability in education. Legislators and other policy makers look for the quickest and cheapest way to provide information on school performance to the taxpayers and voters. As in most states, Illinois relies on a uniform administration of a paper-and-pencil test to provide that information. This is certainly quick and relatively cheap. However, the medium at best can only sample the Illinois Learning Standards, leaving a great deal of learning progress unmeasured and unreported to the public. As many scholars have noted, elected officials want to see results within their terms of office (two or four years), when in reality, we know that real and lasting educational improvements often take a decade or more. When results do not improve within a year, the entire improvement system comes under fire.

- **The general public has a great concern for accountability in education.**

Stemming from the first problem, another arises. When district and school leaders do not fully understand the relative and distinct roles of classroom, district, and state assessments, they tend to focus in on the one perceived to be most removed from the daily life of students, that is, the state assessment. Instead of flowing smoothly into a system in which students are taught the standards content every day, and various assessments corroborate their understanding, the state test is perceived to be the "real" bottom line. And even though that test only samples the Learning Standards because of time, money, and design constraints, school personnel react by "teaching to the test." Then this practice comes under criticism, and rightfully so, by those who realize the true depth and breadth of the Learning Standards. The result? Additional artificial restrictions on state testing, such as a lifetime time limit of twenty-five state testing hours per student. Such restrictions only magnify the problem, and the cycle begins again.

Unless and until more educators and more taxpayers and voters understand the structure and mechanisms of standards-led teaching and learning, and the role of classroom, district, and state assessments within that system, these policy problems will persist. Some solutions, which the Illinois State Board of Education are trying, include the massive state trainings described earlier, access to sample test questions, and scoring on our website, and informational materials for legislators, educators, parents, and students. The new Illinois School Improvement Website (ilsi.isbe.net) shows the public how state test data combined with selected demographics (school size, poverty levels, geographic location, racial characteristics, etc.) can provide a clearer window into school performance and improvement.

**Embedded Assessment –
Richard J. Shavelson**

In this session, a distinction was made between formative and summative assessment. The former is carried out to improve teaching and student learning, while the latter gauges accomplishments at the end of some course of study. Often there is a mismatch between the two, and teachers (and other educators) are caught in the middle. These distinctions were explored and discussion focused on formative assessment that includes embedded assessments and their role in formative assessment.

**Science Performance Assessments: Inquiry and Technological Design –
William F. Fraccaro**

The Illinois Science Performance Standards project was explained, focusing on Illinois Learning Standards 11A and 11B including performance descriptors, performance and classroom assessments, and exemplars of student work indicating meets and exceeds levels.

Science Performance Assessments: Science, Technology, Society, and Accepted Practices in Science – Russ Watson

The Illinois Science Performance Standards project was explained focusing on Illinois Science Standards 13A and 13B including performance descriptors, performance and classroom assessments, and exemplars of student work indicating meets and exceeds levels.

Science Performance Assessments: Content Areas – Trudi Coutts and Larry Cwik

The Illinois Science Performance Standards project was explained including performance descriptors, performance and classroom assessments, and exemplars of student work indicating meets and exceeds levels.

**Assessment of Classroom-Based Inquiry –
Donald Dosch and Susan Styer**

The presenters offered a model of classroom-based inquiry and how it has been assessed. Both teacher and student perceptions were discussed. Session participants were invited to share their plans for or experiences with inquiry and its assessment in their school settings.

Assessment in Problem-Based Learning – John Thompson

Assessment is embedded in the problem-based learning experience in order to not only evaluate student learning, but also to identify areas where further instruction is needed. Various forms of ongoing, authentic assessment aimed at deepening student understanding of concepts were discussed.

Break-Out Sessions

Scientific Inquiries: IMSA's Core Science Program – David Workman

Scientific Inquiries has been designed using the ideas in Wiggins' and McTighe's book, *Understanding by Design*. Examples of design assessments developed were shared along with an evaluation of their effectiveness. Other examples of assessments used within the program were given along with an interpretation of their usefulness in getting at student understanding.

Illinois Virtual High School: On-Line Assessment of Science – David Barr and John Eggebrecht.

The Illinois Virtual High School has been offering courses over the past two years. Of special interest are the ways that science can be taught, but more problematic is on-line science assessment. Innovative ways of assessing on-line science learning were shared and discussed.

Inquiry and Problem-Solving: Meaning-Making in Mathematics and Science – Tonda Hager

Traditionally, inquiry is regarded as being in the domain of science, while problem solving is regarded as being in the domain of mathematics. Illinois and national learning standards in mathematics and science are real-world applications of science and mathematics challenging this artificial distinction. This session explored the implications of inquiry and problem solving for assessment in the two disciplines.

A Model for Assessing Student Presentations – Edwin Goebel

In this session, a model was presented for assessing student presentations in general and pathogenic microbiology. This information stimulated discussion about assessment methods of in-class oral presentations in a wide variety of courses. Discussion focused on sharing methods that "work" and also methods that do not.

Each panelist held a question and answer session based on the morning panel discussion.

Summary

The conversations of the day indicated a general dissatisfaction with the current state of educational assessment. While the dialogue identified a myriad of flaws in current approaches to assessment, it also identified exemplary practices and promising trends. What might the future of assessment look like? While this is impossible to predict, the image that emerged from this Great Minds Dialogue suggests the positive advancements listed below.

THE CURRENT STATE OF ASSESSMENT	THE FUTURE OF ASSESSMENT
• That which is most easily measured, rather than that which is most valued, is what tends to be emphasized on tests.	• Assessment systems will be developed that better represent science as expressed in the National Science Education Standard, and especially deep conceptual understanding and scientific inquiry.
• The form of large-scale assessments (i.e., printed groupings of forced-response and short answer items) tends to be driven by cost constraints in design, production, administration, scoring, and reporting.	• The technology of assessment, and its availability, will progress—allowing more flexible, intelligent, and adaptive methods of assessment.
• The multiple formative and summative purposes of assessment tend to be addressed via distinctly independent means. For example, classroom-based assessments (for "grades") are rarely designed as evidence of program, school, or district performance. Likewise, large-scale tests (for "accountability") tend to provide little detail about the accomplishments of individual children, classrooms, or schools.	• By carefully and diligently aligning both curriculum and assessments to specific learning expectations defined in the Standards, there will be increased coherence in assessment practices, resulting in a more integrated system that more fully informs children, teachers, parents, and school administrators.
• State and district level assessments tend to be administered at "benchmark grades" (i.e., grades 4, 8 and 11).	• Assessments will be administered annually so that the progress of grade level cohorts can be followed over time.
• Tests are administered under "standard" conditions in which the curriculum is suspended and testing is timed. In reality, conditions of testing (i.e., temperature, lighting, seating, preparation) vary widely and children respond differentially.	• Academic standards and improvements in assessment and curriculum technologies will result in increased use of assessment tasks that are seamlessly embedded in curriculum and instruction.
• State assessments are often used to identify failing schools and to sort the others by rank ordering group average scores.	• Assessments will identify the specific strengths and deficiencies of each school in reference to specific learning expectations defined in the Standards.

How will we get from the current state to this more promising future? We heard, in this dialogue, strident messages of the need to truly ground the *enacted* and the *assessed* curriculum in the expectations for learning defined by the National Science Education Standards. We also heard of the great need for the advancement of assessment technologies and the *science* of assessment. Alberts recognized that, due to the prevalence of "local control" in this country's educational system, there are numerous assessment

Summary

"experiments" currently happening throughout the nation. He called for a more substantial and scientific effort to identify and develop more advanced assessment systems that promote "the type of science education outlined in the *National Science Education Standards*." This view seems to be shared by many professional educators and leaders. Recently, the National Research Council released its report *Scientific Research in Education*, describing the current status of educational research, how it arrived at this point, and the criteria for research-driven advancements. The relevance of this document to assessment in education is illustrated in the following quotation.

> The recorded history of testing is over four millennia old (Du Bois, 1970); by the middle of the nineteenth century, written examinations were used in Europe and in the United States for such high-stakes purposes as awarding degrees, government posts, and licenses in law, teaching, and medicine. Today, the nation relies heavily on tests to assess students' achievement, reading comprehension, motivation, self-concept, political attitudes, career aspirations, and the like. The evolution of the science of educational testing, similar in many ways to the progress in genetics, follows a long line of work in educational psychology, psychometrics, and related fields dated back to the late 1800s … Steady but contested and nonlinear progress has been made since the early days… (NRC, 2002; pg. 33).

Perhaps a reasonable conclusion to make of this Great Minds Dialogue is to acknowledge that we find ourselves in a time in which the science and technologies of testing and assessment are rapidly evolving. We might take some encouragement that the current high level of professional and public attention should even accelerate this evolution. However, as examples in this dialogue illustrate, we must also be highly mindful of the potential for error, and also real harm, as assessment "experiments" are tried.

Jumping-Off Points

This annotated list of web sites is designed to provide more in-depth information into some of the assessment points and practices that were raised by the panelists and participants during the dialogue. While it is not exhaustive, they are intended to serve as “jumping-off points” to various topics for those desiring further information.

Achieve, Inc.

http://www.achieve.org/

Achieve was established as a result of the 1996 National Education Summit as an independent, bipartisan, and not-for-profit organization to inform and accelerate standards-based reform.

Consortium for Policy Research in Education (CPRE)

http://www.cpre.org/
http://www.cpre.org/Publications/Publications_Accountability.htm
http://www.cpre.org/Publications/rr46.pdf

CPRE is one of the organizations that has studied assessment and accountability trends over the fifty states. These URLs link to “accountability profiles” reports.

The Council of Chief State School Officers (CCSSO)

http://www.ccsso.org/AccountabilityResources.html

This site provides access to helpful reports including the perspectives of state education officers who are working very intensely on assessment issues. Links to other resources are also provided.

Council of the Great City Schools

http://www.cgcs.org/
http://www.cgcs.org/taskforce/achievegap3.html

Look to this site to learn about the particular effects of testing in the large urban centers, especially the persistence of achievement gaps among demographic subgroups.

CRESST: National Center for Research on Evaluation, Standards, and Student Testing

http://cresst96.cse.ucla.edu/index.htm
http://cresst96.cse.ucla.edu/CRESST/Newsletters/polbrf54.pdf

CRESST is a national research center for research on standards, assessment and accountability issues. The second URL links to an interesting policy brief entitled Standards for Educational Accountability Systems.

Educational Testing Service (ETS)

http://www.ets.org/
http://www.ets.org/fairness/download.html

This is the site of the “world’s largest private educational testing and measurement organization.” The second URL links to the *ETS Standards for Quality and Fairness*.

Education Week on the web

http://www.edweek.org/sreports/qc01/
http://www.edweek.org/context/topics/issuespage.cfm?id=41
http://www.edweek.org/context/states/stateinfo.cfm?stateabbrv=il

Education Week maintains current and comprehensive coverage on trends in assessment including special reports such as the annual *Quality Counts* series. The first URL will take you to the 2001 report on standards and testing. You will also find assessment links at the second URL and issues specific to Illinois at the third URL.

ERICae.net

http://ericae.net/

The ERICae.net site strives to provide “balanced information concerning educational assessment, evaluation and research methodology.” This site provides rapid access to current news and trends in assessment and a popular “test locator” search engine.

FairTest: The National Center for Fair & Open Testing

http://www.fairtest.org/

This is an advocacy organization described as “working to end the abuses, misuses and flaws of standardized testing and ensure that evaluation of students and workers is fair, open, and educationally sound.”

Illinois State Board of Education

http://www.isbe.net
http://.ilsi.isbe.net

This site provides many resources for Illinois educators and contains the Illinois school improvement web site.

Jumping-Off Points

McREL: Mid-continent Research for Education and Learning
http://www.mcrel.org/
http://www.mcrel.org/products/assessment/
This site provides one of the most comprehensive presentations of issues and resources for assessment, ranging from the classroom to large-scale testing.

National Academies of Science
http://www. nas.edu/
The National Academies provide many full text documents and reports free, on-line. Reports can be found under "education" on the home page.

National Academy Press
http://books.nap.edu/v3/makepage.phtml?val1=subject&val2=ed
Relevant texts published by the National Research Council that were cited during this Great Minds Dialogue (see the list of references) can be quickly located by connecting to this site and searching the page for "assessment" or "standards".

National Assessment of Educational Progress (NAEP)
http://nces.ed.gov/nationsreportcard/
NAEP, also known as "the nation's report card," is a regular national sample study of achievement and contextual factors in the principal content areas, including mathematics and science. Students are tested in grades 4, 8 and 12.

National Public Radio (NPR)
http://search.npr.org/cf/cmn/cmnpd01fm.cfm?PrgDate=03/21/2002&PrgID=5
This series of dialogs on National Public Radio's *Talk of the Nation* gives provocative insights into standards-based reform with a focus on assessment, including commentary on the *No Child Left Behind* act.

New York Times' series *None of the Above: The Test Industry's Failures*
http://www.nytimes.com/learning/general/specials/testing/
This site provides information about the four largest testing companies and examples of some recent problems and costly errors.

OECD Programme for International Student Assessment (PISA)
http://www.pisa.oecd.org/index.htm
PISA is an international study of 15-year-olds in principal industrialized countries. This age group is selected because they are near the end of compulsory education. The assessments are focused on the knowledge and skills "that are essential for full participation in society." Performance is assessed in reading, mathematical literacy, and scientific literacy. In addition, the study explores factors of the home and school that influence learning. Policy implications are considered.

PBS *Frontline* documentary: *Testing our Schools*
http://www.pbs.org/wgbh/pages/frontline/shows/schools/
This is an excellent site that explores multiple perspectives about the realities of testing and assessment.

Project 2061
http://www.project2061.org/
http://www.project2061.org/tools/bluepol/blpframe.htm
In addition to the curriculum evaluation work described in this report, Project 2061 provides online access to an excellent chapter on assessment from the book *Blueprints for Science Literacy*.

Third International Mathematics and Science Study (TIMMS)
http://ustimss.msu.edu/
http://nces.ed.gov/TIMSS/
http://isc.bc.edu/timss1999benchmark.html
TIMSS, and the more recent Benchmarking Study (a.k.a. "repeat") TIMSS-R, is an extraordinarily comprehensive assessment of mathematics and science education.

U.S. Department of Education
http://www.ed.gov/
A plethora of resources and reports are available from this site. Be certain to check the "Accountability" link under the "Education Resources" section.

Resources and References

Alberts, B. (1998). Reinvigorating science education in the U.S.: The importance of the appropriate assessments. American Association for Higher Education, Cincinnati.

American Association for the Advancement of Science (1990). This Year in School Science 1990: Assessment in the Service of Instruction. Washington, DC, AAAS Publication.

American Federation of Teachers (2001). Making Standards Matter 2001: A fifty-state report on efforts to implement a standards-based system. Washington, DC, Author.

Atkin, J. M., P. Black, et al. (2001). Classroom assessment and the National Science Education Standards. Washington DC, National Academy Press.

Bransford, J.D., A.L. Brown, R.R. Cocking (eds.) (2002). How people learn: brain, mind, experience, and school. Washington, D.C., National Academy Press.

Feuer, M. J., National Research Council (U.S.). Committee on Equivalency and Linkage of Educational Tests, et al. (1999). Uncommon measures: equivalence and linkage among educational tests. Washington, D.C., National Academy Press.

Goertz, M. E., M. C. Duffy, et al. (2001). Assessment and accountability systems in the 50 states, 1999-2000. Philadelphia PA, Consortium for Policy Research in Education, University of Pennsylvania.

Grissmer, D. W. and Educational Resources Information Center (U.S.) (1998). Education productivity. Washington DC, NEKIA Communications : U.S. Dept. of Education Office of Educational Research and Improvement Educational Resources Information Center.

Herman, J. L. and R. L. Linn (1997). Standards-Led Assessment: Technical and Policy Issues in Measuring School Performance. Los Angeles, CRESST: 31.

Heubert, J. P., R. M. Hauser, et al. (1999). High stakes: testing for tracking, promotion, and graduation. Washington, D.C., National Academy Press.

Koretz, D. M., S. I. Barron, et al. (1998). The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS). Santa Monica CA, Rand.

Merrow, J. (2001). Choosing excellence: "good enough" schools are not good enough. Lanham Md., Scarecrow Press.

National Research Council (U.S.). Committee on Scientific Principles for Education Research., R. J. Shavelson, et al. (2002). Scientific research in education. Washington, DC, National Academy Press.

National Research Council (U.S.) (1996). National Science Education Standards: observe, interact, change, learn. Washington DC, National Academy Press.

Resources and References

National Research Council (U.S.). Committee on Embedding Common Test Items in State and District Assessments, D. M. Koretz, et al. (1999). Embedding questions: the pursuit of a common measure in uncommon tests. Washington, DC, National Academy Press.

National Research Council (U.S.). Committee on Science Education K-12, National Research Council (U.S.). Mathematical Sciences Education Board, et al. (1999). Global perspectives for local action: using TIMSS to improve U.S. mathematics and science education. Washington D.C., National Academy Press.

National Research Council (U.S.). Committee on Title I Testing and Assessment., R. Rothman, et al. (1999). Testing, teaching, and learning: a guide for states and school districts. Washington, D.C., National Academy Press.

National Research Council (U.S.). (2000). Inquiry and the national education standards. Washington, D.C., National Academy Press.

National Research Council (U.S.). (1996).Science education standards. Washington, D.C., National Academy Press.

Project 2061 (American Association for the Advancement of Science) (1993). Benchmarks for science literacy. New York Oxford University Press.

Pellegrino, J. W., L. R. Jones, et al. (1999). Grading the nation's report card: evaluating NAEP and transforming the assessment of educational progress. Washington, D.C., National Academy Press.

United States. Office of Educational Research and Improvement (2002). Highlights from the Third International Mathematics and Science Study-Repeat (TIMSS-R). Washington, DC, U.S. Department of Education.

Webb, N. L. (1997). Research Monograph no. 6: Criteria for Alignment Expectations and Assessments in Mathematics and Science Education. Washington, DC, National Institute for Science Education, Council of Chief State School Officers: ii-39.

Wiggins, G. P. and J. McTighe (1998). Understanding by design. Alexandria, Va., Association for Supervision and Curriculum Development.

List of Participants

Randall K. Backe	SRA/McGraw-Hill
Marjorie G. Bardeen	Fermilab
Diane Bearrick	St. Charles North High
Nancy Benson	Marquardt Middle School
Julia L. Billsten	Nicholson Elementary School
Brandon Birmingham	Prairie Jr. High
Susan Bisinger	IMSA
Warren B. Bjork	Glenbrook South High School
Jane M. Bock	St. Joan of Arc
Gloria-Jean Bogal	Cathedral of St. Raymond
David J. Bonnette	Riverside School District 96
Bob Brazzle	IMSA
Gary A. Brehm	Marian Central High School
Caylee L. Bruce	Palos South Middle School
Bob Burtch	Batavia Middle School
Marjorie Cave	DuPage Regional Office of Ed
Sheila Chilton-Davis	Nazareth Academy
John W. Citek	B J Ward Middle School
Kathryn M. Cunningham	Prairie Jr. High
Raymond J. Dagenais	IMSA
Susan Dahl	Fermilab
Dennis M. Delfert	College of Lake County
Steven E. Deloriea	Marian Central High School
Mary K. Denny	Tefft Middle School
Richard Dods	IMSA
Dana H. Dodson	Valley View School Dist
Don Dosch	IMSA
Pamela S. Duncan	Mundelein High School
Jennifer Eagleton	
John Eggebrecht	IMSA
Kevin Farrell	Marian Central High School
Elizabeth Fisher	B J Ward Middle School
Maria Fontana	
Joe W. Gale	Park Jr. High
Deborah A. Gartshore	Larkin High School
James Gerry	IMSA
Beth Giglio	Stevenson High School
Tricia Gikas	Carl Sandburg High School
Rachel Gladney	Elmwood Park High School
Ed Goebel	IMSA
Janis Goetz	Larkin High School
David Greenwood	Mundelein High School
Judith M. Grossman	Lemont High School
Karin J. Harridge	Highland Upper Grade Center
Katherine A. Hartnett	Thompson Middle School
Mary K. Hemmelman	St. Gilbert Catholic School
Diane Hinterlong	IMSA
Mark Horrell	IMSA
Caroline A. Humes	Adlai E. Stevenson High School
Rita M. Januszyk	Gower West Elementary School
Janis L. Jasper	Marquardt Middle School
Carol D. Johnson	LaGrange School District 102
Thomasina B. Johnson	Naperville North High
Thomas Jordan	Fermilab
Chris Kawa	IMSA
Janet S. Keating	Hinsdale Central High School
Kimberly M. King	Jane Addams Middle School
Lucille E. Kitt	Haines Middle School
Andrea Klingler	St. Thomas More School
Joan M. Kowalczyk	Queen of Peace High School
Ona Kozar	Field Middle School
Lisa M. Krason	Palos South Middle School
David Krodel	St. Charles North High
Anthony Kroll	Mundelein High School
Peggy Kubiak	Cook County School Dist #154
Timothy P. Kulak	West Aurora H S
Branson Lawrence	IMSA
Leon Lederman	IMSA
Timothy R. Leffler	Lemont High School
Dennis D. Lehman	Harold Washington College
Deann R. Liles	Marian Central High School
Mary Lou Lipscomb	IMSA
Deborah L. Lojkutz	Joliet West HS
Kathy M. Lovelace	Lake Park HS
Kathleen D. Ludwig	McGraw Hill
Vincent F. Malek	Marian Central High School
Sharon L. Malito	Palos South Middle School
Stephanie Pace Marshall	IMSA
Ann McCarty	Northbrook School District 27
Edwin A. Metzl	Chicago Public Schools
Gabi Mihalcea	IMSA
Daryl J. Mortensen	Downers Grove South H.S.
Laura Nickerson	IMSA
Tom V. O'Hagan	Marian Central High School
Earle Z. Olson	Downers Grove North H. S.
Douglas Overton	St. Charles North High
Michael Palmisano	IMSA
Korri J. Patterson	Thompson Middle School
Ron Pine	IMSA
Gwen Pollock	IL State Board of Education
Heather J. Pusich	Field Middle School
Dick Raab	Jefferson Jr. High
Sheryll A. Renken	Indian Prairie SD 204
John A. Rhodes	Elm Middle School
James N. Roberts	Riverdale High School
Steven Rogg	IMSA
Caryn E. Rollin	B J Ward Middle School
Patricia A. Salwach	Teft Middle School
Darryl Samborski	School District 15
Mark A. Sandusky	Palos South Middle School
Judith A. Scheppler	IMSA
Jeffrey Schleff	Woodoaks Jr. High
David G. Schultz	Oak Park River Forest HS
Rita L. Slattery	Av Martinez Middle School
James E. Slouf	Downers Grove South HS
Gina Slusinski	Queen of Peace High School
Patrick J. Somers	Libertyville High School
Melissa Staude	Larkin High School
Sue Styer	IMSA
Karen A. Swacker	Wolcott School
John Thompson	IMSA
Meda K. Thompson	Valley View School District
Guy Todnem	IMSA
Joe Traina	IMSA
Joel M. Vickers	Palos South Middle School
James W. Vokac	Maine Est
Jennifer A. Wagner	Bolingbrook
Edwina H. Wallace	Tefft Middle School
Josie Wallmuth	IMSA
Brian Wegley	North High School
Jolene M. Weir	Haines Middle School
Lisa J. West	Frankfort District 157-C
Carol L. Widegren	Lincoln Park High School
Rick A. Wiltshire	CHSD 218
Ann Wohlberg	IMSA
David Workman	IMSA
Jan Wright	Batavia Public Schools
Douglas Zimmer	Wheaton Warrenville South

Illinois Mathematics and Science Academy
1500 W. Sullivan Road
Aurora, Illinois 60506-1000
630-907-5956
http://www.imsa.edu
Email: greatminds@imsa.edu

The IMSA Great Minds Program is made possible by donors to the IMSA Fund for Advancement of Education, grants from the U.S. Department of Education, and support from the State of Illinois. Major gift commitments to the IMSA Fund have been made by American Marketing Systems, Inc./John and Carol Berger, Hansen-Furnas Foundation, Harris Family Foundation, and Tellabs Foundation.