Moving to the Next Generation of Standards for Science: Building on Recent Practices

Joan L. Herman
National Center for Research on Evaluation, Standards and Student Testing (CRESST)
University of California, Los Angeles

Globalization, rapid change, the explosion and specialization of knowledge, the transcendence of information services and knowledge management over the provision of material goods and services, the transformation of the nature of work and social relationships engendered by continual advances in technology, the demand for so-called 21st century skills, the role of science in the future of the world. Contrast these expectations with the findings in America's Perfect Storm (Kirsch et al., 2007), the meager proportions of students who are proficient in science based on NAEP, the disappointing performance of United States students performance on international comparisons, and the scant pipeline of students pursuing careers in math and science. Add to it advances in the theory and practices of learning and assessment. The combination provides a powerful rationale for rethinking the science competencies and dispositions students need to develop in school and the drivers that can help them to get there. Current federal policy both underscores and accelerates the urgency: American Recovery and Reinvestment Act (ARRA, 2009) and its Race to the Top sequel makes clear that standards-based reform remains a key, federal strategy for leveraging school improvement in the United States and assuring that students develop the knowledge and skills they need for future success. As part of ARRA assurances, states needed to commit to "making progress toward rigorous college- and career-ready standards and high-quality assessments that are valid and reliable for all students," which presumably provides a substantial part of the foundation for the other three reform assurances, those of improving teacher effectiveness, establishing longitudinal data systems to inform educational decision making, and providing necessary supports and interventions to schools identified for corrective action or restructuring. *Race to the Top* furthermore require a comprehensive approach to all four reform areas and invite an emphasis on STEM, as well as the coordination and vertical alignment of learning expectations pre-K to 20.

Lessons Learned

This paper argues that a new generation of science standards must be built on lessons learned from the current practice and on recent examples of standards-development methodology. In the sections that follow, I first briefly contrast assumptions about the role of standards in improving learning with research findings on actual effects and use those to highlight essential features for a new generation of science standards. I then describe recent, promising efforts to develop standards in science and other areas, including the NAEP 2009 Science Assessment Framework, the Advanced Placement Redesign, and the recently (unofficially) released Common Core State Standards, and derive from them promising practices for a national effort to define science standards. I end by considering validation issues, i.e., the kinds and claims and evidence one would want to collect to demonstrate that national science standards were achieving their intended purposes.

Role of Standards in Improving Learning

The significant role that standards and assessment can play in establishing and molding new expectations for learning is well documented in research world-wide. The basic idea: (1) Be clear on expectations by establishing standards (2) develop high visibility tests based on the standards; (3) use the test to communicate what is expected, to hold relevant stakeholders accountable for teaching and learning the standards and to provide data to inform needed improvements. Such standards-based tests provide technical evidence for judging performance to serve a variety of decision-making purposes (accountability, selection, placement, evaluation, diagnosis, improvement) but the very existence of the test and the attention it engenders also carries important social, motivational and political consequences.

Research shows the power of some part of the operant model but suggests that it is the test rather than the underlying standards that exert the most significant impacts (see, for example, Herman, in press): High visibility tests serve to focus priorities for curriculum and instruction and tend to drive out what is not tested. Teachers tend to model the pedagogy exemplified in high visibility tests and to mirror the test formats and problem types in instruction. Textbook and other materials publishers modify their products or create new ones to address what is tested, serving as another mechanism for further focusing curriculum and communicating to teachers and students what is needed. These collective findings mean that the nature of tests and other assessments are of signal importance, and, depending on their nature, may serve to encourage transmission- type teaching and a performance rather than mastery orientations to learning (See, for example, Shepard, 2005).

That curriculum, teaching and learning likely emphasize the test rather than the underlying standards emerges from a number of factors, which are cause for multiple concerns. In many states, the standards are vague and do not well communicate to educators, students or test developers what is intended. Absent are clear delineations of content or cognitive demand expectations. That performance standards – i.e., the relationship between score and assigned proficiency level -- are routinely created at the end of a test development and administration process rather than informing that process means that that the relationship between proficiency/achievement levels and knowledge and skill development goals is largely opaque. Absent clearly delineated learning targets, it is difficult for educators to fully understand for what they are being held accountable and thus tend to glean what they can from test content.

Even given clarity, standards in many states evidence other problems that may unintentionally encourage teachers to a focus on the tests. State documents tend to lay out an overwhelming array of standards that surpass available school time for teachers and students to achieve them. Expectations tend to be “a mile wide and an inch deep’ (Schmidt, et al., 2005) , discouraging teaching and learning for understanding. Often lacking a coherent sequence of development within or across grades, or ties to organizing principles of the field, standards docs too often seemingly lay out adhoc lists of content expectations that miss important educative opportunities for teachers and students – for example, the learning value of organizing principles (NRC, 2000). Faced with a bewildering array of standards and strong accountability demands, educators may have little choice but to focus on what is tested (See Wilson & Berenthal, 2005) for a more complete analysis of problems with current science standards).

The set of circumstances that encourages educators to focus on what’s tested functionally leaves decisions about what is taught in the hands of item writers and test developers. This is particularly problematic in that studies of the alignment of standards and tests show that typical state tests emphasize lower level knowledge and skills inherent in state standards at the expense of complex thinking, problem solving and other 21st century competencies (See, for example, Webb, 1999).

Clearly, current standards-based systems are not getting students to where they need to be, as the NAEP and international results mentioned earlier attest. In contrast, they are producing students who are not prepared for college or for the demands of the workplace (see, for example, Conley2007; Schneider, 2009) These multiple shortcomings of current standards have led to prevailing mantra that standards must be “fewer, clearer, higher (FCH),” meaning in general that standards should:

- Define an *essential core set* of academic competencies that students can feasibly achieve and need for post-secondary access;
- Be sufficiently clear to guide the development of assessment to support accountability and improvement for students, educators, administrators and the system as a whole; and
- Be sufficiently clear to guide the design and provision of rigorous coursework and engaging teaching and learning opportunities to enable students to achieve such competencies
- Represent the knowledge, skills and competencies that students need to be prepared for success in college and the workplace.
- Be benchmarked to the international standards and directly address the knowledge and skills that will enable students to be successful students of the 21st century. (Gates, 2009)

Clarity is an over-riding essential feature, in that without it, one cannot judge whether one has defined the essential core for post-secondary success or whether this core represents knowledge and skills that are internationally competitive. Clarity clearly is essential as well for a strong foundation for an aligned science education system that can guide teaching and learning. Systems for State Science Assessment defined criteria for achieving clarity of standards as:

- Be clear, detailed and complete;
- Be reasonable in scope
- Be rigorously and scientifically correct
- Have a clear conceptual framework
- Be based on sound models of student learning; and
- Describe performance expectations and identify proficiency levels. (Wilson & Berenthal, 2005, p.62)

Drawing on *Knowing What Student Know* (Pellegrino, Chudowsky & Glaser, 2001), Wilson & Berenthal (2005) also emphasize the need for multilevel assessment systems that can provide coherent, comprehensive, and continuous data to state, district, schools and particularly teachers to improve student learning. They suggest the development of standards and assessment with an eye toward the system you want to create.

Recent Standards Development Methodologies

Recent standards-development projects have attempted to bring more clarity to the standards they develop, including the specification of content and cognitive demand, as well as to help assure that the standards reflect the rigor expected for post secondary success. While the basic approach to standards development is similar across projects – i.e., assemble subject matter experts, have them use their own knowledge and experience with existing standards and relevant research to articulate a new set, vet and improve the standards through feedback from other experts and constituency groups-- each has some unique features and lessons learned. I consider here methodologies used to develop (1) the NAEP 2009 Science Framework; (2) AP Redesign; and (3) Common Core:

NAEP Science Framework (NAGB, 2008)

The NAEP Science Framework development process, initiated and overseen by the National Assessment Government Board, involved hundreds of individuals from across the country, including leading scientists, science educators, and measurement experts. Overall direction and periodic review for the effort was provided by a Steering Committee representing key policy and practice constituencies and national organizations committed to science education. A designated Planning Committee, likewise composed of scientists, K-12 and higher education science educators and assessment specialists, were responsible for practical framework development. The Planning Committee used existing national standards (NAS, 1995; AAAS, 1989), state standards, and international assessment frameworks (Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA)) and associated research to draft the initial framework As the framework progressed, it was vetted in a process of regional hearings and other public forums and was revised based on feedback from these venues. The National Assessment Governing Board also engaged an independent review of the draft and convened public hearings to solicit feedback.

The wealth of content in source materials was condensed into key foundational principles and pervasive understandings in each of Life, Physical and Earth and Space science for NAEP's 2009 Science Assessment. Test content is specified relative to content and cognitive demands, concentrating on topical themes that transverse grades 4, 8, and 12. Framework developers intended a focus on students' conceptual understanding and ability to apply science concepts, principles, laws and theories and also incorporated components of scientific inquiry and technological design. Intended emphases for the NAEP assessments are specified at each grade level in terms of content, science practices, item types and item distributions across content, practices and type:

Science content. Drawing on key facts, concepts, principles, laws and theories that cross grade spans in each discipline, central principles are specified for each discipline in an intended progression of complexity of understanding from grade 4 to grades 8 and 12. Content statements at each grade level, in contrast to the prior framework's topic lists, are articulated as propositions that are intended to express science principles that represent the consensus of the scientific community:

<u>Physical science</u>: matter, energy, and motion, which in turn are described in terms of subdomains:

- **Properties of Matter:** Physical properties common to all objects and substances and physical properties common to solids, liquids, and gases (4); chemical properties, particulate nature of matter, and the Periodic Table of the Elements (8); characteristics of subatomic particles and atomic structure (12). (p 33)
- **Changes in Matter:** Changes of state (4); physical and chemical changes and conservation of mass (8); particulate nature of matter, unique physical characteristics of water, and changes at the atomic and molecular level during chemical changes (12). (p.34)
- **Forms of Energy:** Examples of forms of energy (4); kinetic energy, potential energy, and light energy from the Sun (8); nuclear energy and waves (12). (p. 35)
- **Energy Transfer and Conservation:** Electrical circuits (4); energy transfer and conservation of energy (8); translational, rotational, and vibrational energy of atoms and molecules, and chemical and nuclear reactions (12). (p. 36)
- **Motion at the Macroscopic Level:** Descriptions of position and motion (4); speed as a quantitative description of motion and graphical representations of speed (8); velocity and acceleration as quantitative descriptions of motion and the representation of linear velocity and acceleration in tables and graphs (12). (p.37)
- **Forces Affecting Motion:** The association of changes in motion with forces and the association of objects falling toward Earth with gravitational force (4); qualitative descriptions of magnitude and direction as characteristics of forces, addition of forces, contact forces, forces that act at a distance, and net force on an object and its relationship to the object's motion (8); quantitative descriptions of universal gravitational and electric forces, and relationships among force, mass, and acceleration (12). (p. 38)

<u>Life science</u>: structures and functions of living systems and changes in living

systems

- **Organization and Development:** Basic needs of organisms (4), levels of organization of living systems (8) the chemical basis of living systems (12). (p.45)
- **Matter and Energy Transformations:** The basic needs of organisms for growth (4), the role of carbon compounds in growth and metabolism (8), the chemical basis of matter and energy transformation in living systems (12). (p.46)
- **Interdependence:** The interdependence of organisms (4), specific types of interdependence (8), consequences of interdependence (12). (p.47)
- **Interdependence:** The interdependence of organisms (4), specific types of interdependence (8), consequences of interdependence (12). (p.48)
- **Evolution and Diversity:** Differences and adaptations of organisms (4), preferential survival and relatedness of organisms (8), the mechanisms of evolutionary change and the history of life on Earth (12). (p.49)

Earth and space sciences: Earth in space and time, Earth structures, and Earth systems

- **Objects in the Universe:** Patterns in the sky (4), a model of the solar system (8), a vision of the universe (12). (p.57)
- **History of Earth:** Evidence of change (4), estimating the timing and sequence of geologic events (8), theories about Earth's history (12). (p. 58)
- **Properties of Earth Materials:** Natural and manmade materials (4), soil analysis and layers of the atmosphere (8). (p.59)
- **Tectonics:** The basics of tectonic theory and Earth magnetism (8), the physical mechanism that drives tectonics and its supporting evidence (12). (p.60)
- **Energy in Earth Systems:** The role of the Sun (4), the Sun's observable effects (8), internal and external sources of energy in Earth systems (12). (p. 61)
- **Climate and Weather:** Local weather (4), global weather patterns (8), systems that influence climate (12). (p.61)
- **Biogeochemical Cycles:** Uses of Earth resources (4), natural and human-induced changes in Earth materials and systems (8), biogeochemical cycles in Earth systems (12). (p. 62)

Science practices. Four inter-related science practices' dimensions define the performance expectations for the specified content:

- Identifying scientific principles: integral to all other practices, this category includes students' ability to recognize, recall, define, relate, and represent basic science principles specified in the content statements.

- Using scientific principles: ability to use principles to explain observations; make predictions; suggest examples, propose, and evaluate alternative explanations.

- Using scientific inquiry (recognized as addressing selected components only): ability to design or critique investigations, conduct investigations using appropriate tools and techniques, analyze data patters and use evidence to validate or evaluate conclusions and explanations.
- Using technological design: ability to develop or evaluate solutions to practical problems, identify tradeoffs and choose among alternative solutions, apply principles to anticipate effects of design decisions.

Communication is explicit as a cross-cutting expectation that permeates each of the practices. Moreover, the framework lays out "cognitive demands" as another lens through which to view these practices, i.e., "knowing that," "knowing how," "knowing why," and "knowing when and where to apply knowledge" (p.91, see also Shavelson et al., 2005). Note that these practices and cognitive demands are a departure from the processes that were identified in the prior NAEP science framework. In addition, contrary to the prior framework, history and nature of science are incorporated as contexts for assessment items rather than as separate topic categories for assessment.

Item distribution. The framework establishes expectations for the emphasis to be accorded each of the three science disciplines and each of the major practices. The framework also defines and lays out expected distributions for the use of specific item types.

At grade four, each of physical, life and earth/space sciences are to be accorded equal attention. Earth space sciences is then accorded relatively more attention at grade 8 and relatively less attention at grade 12 relative to the other two areas.

Across all grade levels, the Framework accords 60% of available testing time to the practices of Identifying Science Principles and Using Scientific Principles, with the latter gaining in emphasis relative to the former as one moves to grade 8 and then 12. Thirty percent of testing time is allocated to Using Scientific Inquiry and the remaining 10% of available testing time to Using Technological Design.

According equal testing time to selected- and constructed- response items, the Framework lays out the following item types for the main assessment: (p.98):

1. Selected response

— Individual multiple-choice items

2. Constructed response

Short constructed-response items

Extended constructed-response items

Concept-mapping tasks

3. Combination

Item clusters

POE item sets (Predict, Observe, Explain)

Combination items involve a related set of items that may be constructed, selected, or a combination of types.

In addition, the framework specifies the conduct of a special, additional assessment for a subsample of students to engage in hands-on performance tasks and interactive computer tasks. Intended to oncentrate on the assessment of complex thinking and problem-solving, these task types are defined as:(p ix):

Hands-on Performance Tasks

In hands-on performance tasks, students manipulate selected physical objects and try to solve a scientific problem involving the objects. NAEP hands-on performance tasks should provide students with a concrete task (problem) along with equipment and materials. Students should be given the opportunity to determine scientifically justifiable procedures for arriving at a solution. Students' scores should be based on both the solution and the procedures created for carrying out the investigation. Further discussion about hands-on performance tasks can be found in chapter four.

Interactive Computer Tasks

There are four types of interactive computer tasks: (1) information search and analysis, (2) empirical investigation, (3) simulation, and (4) concept maps. Information search and analysis items pose a scientific problem and ask students to query an information database and analyze relevant data to address the problem. Empirical investigation items place hands-on performance tasks on the computer and invite students to design and conduct a study to draw conclusions about a problem. Simulation items model systems (e.g., food webs) and ask students to manipulate variables, and predict and explain resulting changes in the system. Concept map items probe aspects of the structure or organization of students' scientific knowledge by providing concept terms and having students create a logical graphical representation.

Framework developers specified that at least one of each type and no more than four should be included at each grade level.

Communication issues. The Framework developers appear to have taken special pains to communicate their intentions to a broad audience. The document includes examples documenting expectations for content, scientific practices and item types. Special clarification boxes are found throughout the document to enable readers to differentiate concepts and see connections across disciplines – for example, the difference between "Identifying Scientific Principles" and "Using Scientific Principles," the ways in which topics across disciplines relate to common themes and models. Sample items are liberally used throughout the text to illustrate the ways in which content and practice intersect to create performance expectations, how items can be generated and interpreted and provide promising assessment practices that users may model. Expected science content in presented in detailed, cross-grade charts that also allow the reader to see the intended progressions and how the complexity and breadth of student understanding is intended to grow.

Advanced Placement (AP) Redesign (Huff & Plake, in press**)**

While the NAEP Science Framework breaks new ground in a number of areas, e.g., its conceptualization of content as science principles, use of implicit learning progressions, and definitions and demonstrations of performance expectations relative to the intersection of content and practice domains, the AP redesign moves standards development of a new generation of specification and possibilities for alignment. Drawing on Evidence-Centered Design (ECD) principles (Mislevy & Risconscente, 2006) and recent research and theory in science learning, the redesign specifies expectations for each subject assessed by AP relative to the intersection of detailed concept maps underlying enduring disciplinary principles and a cognitive framework specifying sets of science practices that are designed to be common across science disciplines. The effort is particularly unique in pre-specifying content and practice demands in terms of "claims" that define what students should know and be able to do to be classified at a particular achievement level – i.e., specific performance expectations that define the capacity expected for students who attain a score of 5, those that define a score of 4, of 3. It also lays out specific evidentiary requirements for establishing each claims

These specifications and evidentiary requirements are then to be used to generate AP tests as well as to form a strong framework for guiding teaching, materials development and professional development. The intent, at least in part, is to respond to concerns that the breadth of current advanced courses gives short shrift to developing depth of student understanding and ability to apply science (see NRC, 2002). The Redesign thus aims to both limit the breadth of content addressed in AP courses and simultaneously to increase students' engagement with scientific reasoning, inquiry and deep conceptual understanding of disciplinary content. The Redesign in science includes AP courses in biology, chemistry, environmental science, and physics

Structure of the process. The College Board used a highly structured process fo developing and reviewing detailed learning domain analyses for each of the four science disciplines. Commissions appointed for each discipline were charged with the domain analysis, which sought to bring together essential content, reasoning, and inquiry skills with enduring principles to create a map of each learning domain. Each commission was composed of a balance of practicing scientists, university faculty and high school science educators (n==12), who ostensibly represented visionaries with regard to science and science education and experts in both science and science teaching and learning. Some also were experienced AP teachers Commissions met at least four times over approximately 9 months to produce initial review drafts.

Drafts were then reviewed by Peer Review Panels in each discipline, which essentially mirrored the expertise composition of the initial Commissions. The Review panels' charge generally considered the extent to which the learning domains achieved the goals of the redesign and represented modern and accurate perspectives on the discipline.

Review Advisory Panels for each discipline, composed of two Commission members and two members of the Peer Review Panel, then worked to refine the domain specifications based on prior feedback and to further incorporate achievement level claims. These refinements also incorporated a common cognitive framework developed by a commissioned Learning Panel, a group of experts in learning in each of the domains. The framework defines the ways in which students are expected to both acquire and demonstrate their competence in the domain, incorporating reasoning and inquiry skills that also are intended as learning targets. Neither the content nor the cognitive/practices elements of the domain exist in isolation; each requires the other for meaning.

Structure of the "content" domain analysis. Each commission started by defining and agreeing on the major ideas to be addressed by each course – 4-7 major ideas for each -- and then worked in subgroups to define the enduring understandings that are essential to each major idea (called level-two concept) and the more specific concepts (level-three concept) that underlie each enduring understanding. The level-3 concepts provide specificity in defining what does and does not lie within the intended course domain. It is worth noting that the instructional time required to develop meaningful understanding of each level-3 concept was a continuing touchstone for defining a realistic domain for teaching and learning.

For example, one of the major ideas specified in Chemistry is "Changes in matter involve the rearrangement and/or reorganization of atoms and/or the transfer of electrons. Among the enduring understandings thought to support this major idea is that of "Chemical changes are represented by a balanced chemical reaction that identifies the ratios with which reactants react and product form.   And in support of this enduring understanding were supporting understandings such as:

- A chemical change may be represented by a molecular, ionic, or net ionic equation.
- Quantitative information can be derived from stoichiometric calculations that utilize the mole rations from the balanced equations.
- Etc. (Ewing, Packman, Hamen & Clark, 2009, p. 26)

Structure of the "practice" domain analysis. As noted above, the Learning Panel in collaboration with disciplinary experts created a framework to define the practices in which students were to be engaged to acquire and demonstrate competence in science. While the framework operationalizes seven key practices that are intended to apply to all AP science courses, the developers recognize that specific instantiations likely would need to be customized for each discipline. The seven practices include (College Board, 2009):

1. The student can use of representations and models to communicate scientific phenomena and solve scientific problems.
2. The student can use of mathematics appropriately
3. The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.
4. The student can plan and implement data collection strategies in relation to a particular scientific question.

5 The student can perform data analysis and evaluation of evidence

6 The student can work with scientific explanations and theories.

7 The student is able to connect and relate knowledge across various scales, concepts, and representations in and across domains

These are broad categories that are further subdivided into more specific components to provide meaningful targets for instruction and assessment. For example, the category of "use mathematics appropriately" includes specific expectations for such things as students' ability to justify the selection of a mathematical routine to solve problems, ability to apply mathematical routines to quantify natural phenomena, etc. while the category of engagement in scientific questioning includes such skills at the ability to pose and evaluate scientific questions. (Ewing et al., 2009, p. 27)

The Practice domain analysis also includes specification of the types of evidence that could substantiate competence in the specific elements of the practice domain. For example, the ability to apply mathematical routines to quantify natural phenomena includes evidence statements such as:

- Appropriateness of application in new context
- Correctness of mapping of variables and relationships to natural phenomena
- Reasonableness of solution given the context
- Prediction of the dynamic relationships in the natural phenomena
- Precision of values consistent with context. (Ewing et al., 2009, p 27)

As another example, evidence of the ability to connect concepts in and across domains to generalize or extrapolate in and/or across enduring understandings and/or big ideas (an element of the ability to connect and relate knowledge across various scales, concepts, and representations in and across domains) includes such evidence statements as:

- Articulation of content-specific relationships between concepts or phenomena;
- Prediction of how a change in one phenomenon might effect another;
- Comparison of salient features of phenomena that are related
- Etc. (Ewing et al., p. 28).

Domain models for each course. The domain analyses for content and practices/ skills then were used to specify the intended domain for each course. See Figure 1 (from Huff, 2009). Expert panels crossed content and practice components to create specific claims that operationalize competency expectations for each AP achievement level. They also articulated statements of the evidence required to substantiate each claim. Effectively, then, the panels pre-specified the performance standards to be used to classify students at a score level of 3, 4, or 5, the scores that determine whether students qualify for college course credit.

INSERT FIGURE 1 ABOUT HERE

These claims then functionally represent a latent performance continuum, which span and defines the achievement levels. The claims also are the foundation for the assessment framework that that then uses the claims to specify assessment task models and assembly specifications, as summarized in Figure 2, taken from Huff, Steinberg & Matts, 2009)

INSERT FIGURE 2 ABOUT HERE

Common Core State Standards

Sponsored by the Council of Chief State School Officers (CCSSO) in collaboration with the National Governor's Association (NGA), recently, unofficially released Common Core State Standards drafts in English-language arts and mathematics (see respectively www.edweek.org/media/draft_standards_for_reading_writing_communication_7-14-09.pdf and www.edweek.org/media/draftmathstandards-july162009-07.pdf) are noteworthy in any number of respects. These include issues of intent, process – who was involved and over what time period – goals, and attention to literacy in subject matter content.

In the words of its developers, the

> "Common Core State Standards Initiative is a significant and historic opportunity for states to collectively accelerate and drive education reform toward the ultimate goal of all children graduating from high school ready for college, work, and success in the global economy. The initiative will build off of the research and good work states have already done to build and implement high-quality standards. The standards will be research- and evidence-based, aligned with college and work expectations, include rigorous content and skills, and be internationally benchmarked (Common Core Standards Initiative, 2009)."

The standards are explicitly being developed to serve as a foundation for curriculum and instruction, professional development and assessment. Moreover, while not intended as either national standards (states agreeing to the Common Core commit that it will represent at least 85% of their standards) or the basis of a national test, the federal government plans to invest $350 million for state and/or consortia of states to develop new assessments to align with the core.

Development process. In terms of the process of development, the pace to rapid: While both NAEP and AP specifications were initially developed over an 18 month period, the high school Common Core draft development was accomplished in less than six (the official intent to develop national standards was not made public until June 1, 2009!). This initial development, directed at expectations in each subject for college and work readiness at the end of high school, has been conducted by panels composed chiefly of representatives of organizations who have been deeply involved in standards development and the assessment of college readiness, i.e., Achieve, College Board, and ACT, augmented by additional, independent subject matter experts. The initial drafts are being reviewed by independent feedback panels in each subject area, composed of subject matter, and assessment/measurement experts. Subsequent to that review and revision, a validation committee composed of additional experts will review the process and substance of the standards "to ensure they are research- and evidence-based and will validate state adoption of the common standards (CCSI, 2009)." Meanwhile, K-12 standards are being developed by backward chaining from high school expectations, and political support is being garnered, in part, through a National Policy Forum of supporting national organizations, for example, the Alliance for Excellent Education, Business Roundtable, Council of Great City Schools, Hunt Institute, National Education Association, National Association of State Boards of Education, National School Boards' Education.

The official timeline:

- August 2009 – draft of common core state standards for college and career readiness English-language arts and mathematics completed and publicly released by standards development committee.
- September 2009 – college and career readiness standards approved by validation committee
- December, 2009 – K-12 common core state standards in English-language arts and mathematics completed and publicly released.
- January 2010 – K-12 standards approved by validation committee
- Early 2010, states submit timeline and process for adoption of common core state standards in English-language arts and mathematics.

Goals and evidence base. The Common core was launched with the official intent to represent the "Fewer, Clearer, Higher" standards that students need to be prepared for success post- high school graduation. This focus represents a notable effort both to align K-12 education with the post secondary expectations, and to explicitly map back from expectations at high school graduation to specific grade-by-grade K-12 standards. Ostensibly aligned with college and work expectations, the standards also are intended to incorporate higher order skills, abilities to apply knowledge and other 21st century skills as well as to be internationally benchmarked to assure global competitiveness. Rather than being aspirational, as initial standards from the last generation tended be, the Common Core also is intended to be a ambitious but realistic set of competency expectations.

Moving from primary reliance on expert opinion, the Common Core also claims to be *evidence-based*. For example, the mathematics group consulted national reports and recommendations on mathematics and mathematics learning (e.g., Adding it Up, Focal Points, How People Learn, Niss' Quantitative Literacy and mathematical competencies); research on requirements for college readiness, such as that conducted by Achieve (2008), ACT, College Board and David Conley (2007, 2008); career readiness analyses conduct by ACT, Achieve's American Diploma Project , and state studies; and documents laying out expectations and/or curriculum guidelines in countries showing the highest performance in international comparisons, such as Belgium, China, India, Korea, Japan, Finland, and Singapore.

Organization. In mathematics, the unofficial released document is ten mathematical principles, with associated explanations what constitutes a coherent understanding of the each principle, – for example Number, Expressions, Equations, Functions, Modeling. In addition to a statement of t a Coherent Understanding of the Principle means, the standards also describe core concepts and skills that constitute the understanding. Sample tasks and problems are used to illustrate and delimit the range of content expected.

In addition to the core principles, the standards also contain a set of Mathematical Practices that are considered key to success in the workplace, college, and the 21st century. Among these practices are that students:

- care about being precise,
- construct viable arguments,
- make sense of and persevere in solving complex problems,
- look for structure,
- look for and express regularity in repeated reasoning,
- make strategic decisions about the use of technological tools.(p.6)

Similar generally to the NAEP and AP frameworks, the Common Core standards are described relative to both expected content understandings and core practices.

Attention to content area reading, writing, and communication. A final note about the core standards in Reading, Writing, and Communications: they lay out expectations for reading informational texts, such as those in science, and standards for writing, speaking and listening to be ready for college, subject-oriented coursework. Similarly, they speak to Application of the Core in the areas of research and use of media. Clearly these also are issues of importance in science teaching and learning.

Validation of Standards[1]

It appears that there have been more advances in developing standards than in attention to their validation, which typically has relied solely on expert review and feedback. But just as it is possible to incorporate evidence into the standards development process, so too is it reasonable to consider how the validation of standards could be more evidence-based.

What does it mean to validate a set of standards? Validity in common parlance denotes the state of being well grounded or justifiable, of being efficacious or producing intended ends, and/or possessing legal and binding force (Merriam-Webster). For the educational measurement community, the concept denotes evidence of how well a test serves its intended purposes and requires the accumulation of a variety of evidence to make the argument that test scores are appropriate for each proposed use (AERA, APA, & NCME, 1999; Kane, 2004). Drawing from these perspectives, the validation of standards would involve the articulation of the purposes such standards are intended to serve; development of an interpretative argument to establish claims that the standards must satisfy to accomplish their purposes, and finally the development of an evidence base to substantiate the argument and verify the claims.

Purposes and Claims

As noted above, today's standards are to be "fewer, clearer, higher." In identifying constructs for teaching, learning and assessment, then the standards in essence should:

- Define an essential core set of academic competencies (FEW) that students need for post-secondary success and as citizens of the 21st century (HIGH);
- Be sufficiently CLEAR to:
    - i. guide the design and provision of rigorous and engaging coursework and learning opportunities to enable students to achieve such competencies, as well as to assure that teachers are prepared to support such learning
    - ii. Undergird the development of formative and summative assessment systems to support accountability and improvement for students, educators, administrators and the system as a whole; and

---

[1] [1]This section adapted from an unpublished manuscript prepared for the Gates Foundation by Joan Herman and Eva Baker (2009).

- Reflect knowledge, skills and capabilities that will enable students to be internationally competitive in today's global economy (HIGH).

In addition, clearly the standards must be defensible, in terms of access and fairness for all individuals.

With these qualities, it is expected that the standards will be useful and used to foster intended consequences, that is, more rigorous, engaging coursework and learning opportunities for students, particularly for students at risk; students entering college and the workforce more prepared than currently for success; students more successful in college and workforce. While a full scientific investigation of all of these claims may not be feasible, they do suggest the kinds of concerns and evidence that ought to be in the forefront in the standards development and validation process. The following are intended as examples (also see Common Core Standards Criteria, 2009):

"Fewer"

- Represent a powerful and coherent set of essential competencies that students' can accumulate by grade and over the course of their K-12 education
- Represent a coherent, vertical progression of knowledge and skills development within and across grades, where appropriate to the structure of content.
- Are reasonable, while still cognitively demanding, in scope. such that all students can be expected to acquire them to graduate high school.

"Clearer"

- Sufficiently clear and specific to be guide consistent teaching, learning and assessment, including the development of curriculum and instructional materials.
- Are clearly communicated so that intended users understand, uniformly interpret, and are able to use the standards for intended purposes (instruction, assessment, professional development, and support systems)
- Are unambiguously communicated through multiple representations. Common language versions must be supplemented to make expectations explicit, e.g., through provision of glossaries with unambiguous definitions, use of graphical, tabular, or other transparent representations, use of sample tasks or problem types
- Clearly define expected levels of content and cognitive demand, e.g., through explicit definition of eligible problem types, criteria for determining quality of response, expected levels of cognitive complexity, and differentiation (based on content review by content and learning specialist)

"Higher"

- Represent the preparation and competencies students need to be successful in college coursework and/or livable-wage workforce training.

- Incorporate deep conceptual understanding and high levels of cognitive demand, including abilities to apply knowledge, reason, conduct inquiry, communicate.
- Explicitly require transfer (beyond item format) to different situations and conditions,
- Are globally competitive: aligned with, extends standards/expectations in highest performing PISA and IEA countries, e.g., Finland, Korea, Netherlands, Canada, New Zealand, Australia, Singapore, Chinese Taipei, and Hong Kong); aligns/goes beyond PISA's performance expectations (based on expert benchmarking)

"Defensible"

- Meet the criteria of content accuracy, fairness to groups with different language and cultural backgrounds, be susceptible to assessments using a variety of formats, and present cost data if possible for renewing item or task sets, scoring, including people, AI, computer display, monitoring, and reporting of results, if applicable.
- Are instructionally sensitive, if standards are intended to form the basis of an instructional program, to evaluate individuals or institutions. That is, standards represent learning or training goals, rather than the description of a normally distributed trait or ability.

Evidence Base

While it is beyond the scope of this paper to lay out specific study designs for accumulating an evidence base for validating standards, suffice to say that a variety of types of studies and evidence would be needed, including:

- *Content evidence, based on subject matter experts and workforce specialists review, including benchmarking studies comparing the standards against known post secondary expectation and/or other sets of standards thought to represent high standards (e.g., those of internationally high scoring countries, those which underlie international assessments)*
- *Empirical evidence from special studies*, including retrospective analyses of available, existing evidence (e.g., high school and college transcripts; scores on various secondary standardized assessments; SAT/ACT); expert /novice and predictive studies to substantiate the value of specified knowledge and skills and future success, and other empirical studies to substantiate specific validity claims with regard to feasibility and utility.

Summary and Conclusions

Standards-based reform continues to be the central framework underlying state educational policy (Massell, 2008), stimulated at least in part by federal education programs and massive stimulus investments (ARRA, 2009). Yet, policymakers and researchers have become increasingly aware of the shortcomings of current efforts and the shaky foundation that many states' content standards provide for development of coherent programs and practices to improve student learning. Rather that providing a clear roadmap for guiding teaching, learning and assessments, current standards too often feel more like overwhelming, adhoc lists of topics without sufficient regard for either how students learn and develop understanding in academic subjects, how fundamental ideas and understanding may develop over time, and what capabilities students will need to be prepared for college, work and/or to be successful in the 21st century, e.g., the ability to access and apply knowledge, use it to reason, ,conduct inquiry, solve problems, innovate. Absent clear guidance from standards, educators have relied on what is tested to focus curriculum and instruction. This reliance on the test to define curriculum and instruction functions tends to devalue more complex cognitive skills relative to more rote ones.

Recent approaches to developing standards in science, including the NAEP Science Assessment Framework for 2009 and the AP redesign, respond productively to some of these challenges. Both efforts have delimited their domains of interest through the articulation of "big ideas" of the discipline, and in the case of AP, through the delineation of the enduring understandings and specific concepts that are to underlie each idea. Moreover, both efforts define performance expectations in terms of the intersection of specific content and cognitive demands/practices, recognizing that one without the other is meaningless. However, the cognitive demands/practices defined by each vary.

Both efforts also are attentive to learning progressions, the NAEP framework more generally so in terms of how understanding of focal content generally develops from grade 4 to grades 8 and 12, and in the AP context, through more specific attention to the development of explicit performance continua detailing claims that should apply to students scoring at given levels (3, 4, and 5). This aspect of the AP approach also is noteworthy in its attempt build a substantive continuum to underlie each course and thus to pre-specify the substantive meaning of proficiency score values. These specifications should enable AP to develop tests that will explicitly differentiate these substantive meanings, rather than making performance standard-setting (e.g., proficient or not) an after-the-fact judgment call based on individual items and tasks and item performance.

Clearly, the AP redesign is the more highly specified of the two, and provides a potentially interesting model for the "Fewer, Clearer, Higher" standards that are to be the focus of the Common Core State Standards. The AP redesign is explicitly intended as a framework to guide course level teaching and learning, resource development and professional development, as well as the AP assessments. As developers moved from domain analysis to domain models and assessment frameworks, their targets become increasingly more specific and the process seems to promote transparency.

The AP assessment frameworks are not yet available, yet one can see from the NAEP framework some advantage for considering the assessment as part of the standards development process. The NAEP Framework not only establishes expectations for selected response and long and short constructed response items, including hands-on performance tasks, but also provides new and innovative models for assessing science, including technology and simulation-based tasks. Throughout, the document makes extensive use of sample items and tasks to document and clarify the framework's intentions.

The Common Core State Standards draft represents a first attempt to respond to current calls for "Fewer, Clearer, Higher" standards that represent ambitious yet feasible goals for all students' college preparedness and readiness for success in 21st century work. The Common Core is starting with expectations for competency at high school graduation that are aligned with post-secondary demands, in both college and 21st, century, living wage work, naturally aligning K-12 with what comes next for students. After establishing these expectations, the effort is working to backward map these standards to create coherent, grade-by-grade expectations.

Like NAEP and AP, the Common Core defines standards in terms of both required "big ideas" and their constituent understandings/concepts and expected cognitive demands/practices required for content competency. The effort also is notable in its apparent attention to theory and research in learning, available research on the meaning of preparedness – e.g., evidence documenting the capabilities required in college coursework and in the workplace – and rigorous standards – e.g., benchmarking relative to the standards and curriculum expectations in internationally, high scoring countries.

Finally, the watchwords of "Fewer, Clearer, Higher" may be seen as both guiding principles for the development of new standards for science and as the bases for claims that need to be substantiated to validate any standards produced. My bias says that as difficult as it may be to negotiate, "clearer" is the key pre-requisite. To repeat: Without "clearer" there is no way to know whether standards are "fewer" or "higher."

References

AAAS (1989) *Science for all Americans.* Washington, DC:AAAS

College Board (2009) *AP redesign project: Progress report to National Science Foundation.*

College and Career Readiness Standards for Mathematics. (July 16 2009). Draft for Review and Comment. Retrieved at www.edweek.org/media/draftmathstandards-july162009-07.pdf August 4, 2009.

College and Career Reading Standards for Reading, Writing and Communication (2009, nd) Retrieved at www.edweek.org/media/draft_standards_for_reading_writing_communication_7-14-09.pdf August 4, 2009.

Common Core Standards Initiative (2009) Common Core State Standards Initiative

Frequently Asked Questions. Retrieved 8/05/09 from http://www.corestandards.org/FAQ.htm

Conley, D. (2007) *Redefining college readiness.* Eugene, OR: EPIC

Ewing, M., Packman, S., Hamen, C., Clark, A. (2009) Representing the targets of measurement within ECD. Annual meeting of the National Council of Measurement in Education. San Diego, CA.

Herman, J. L. (2008) Accountability and assessment in the service of learning: Is public interest in K-12 education being served? In L. Shepard and K. Ryan (eds).*The future of test-based accountabilit*y. New York: Taylor and Francis

Huff, K. & Plake, B. (in press). The AP Redesign. *Applied Measurement in Education.*

Huff, K; Steinberg, L; & Matts, T. (2009) The Promises and Challenges of implementing evidence-centered design in large-scale assessment. Annual meeting of the National Council of Measurement in Education. San Diego, CA.

Kirsch, I., Braun, H., Yamanoto, K. & Sum, A. (2007) ***America's Perfect Storm: Three Forces Changing Our Nation's Future.*** Princeton, NJ: ETS

Massell, D. (2008). *The current status and role of standards-based reform in the United States.* Paper prepared for the National Research Council Workshop on Assessing the Role of K-12 Academic Standards in the States.

Mislevy, R. & Risconscente, M. (2006) Evidence centered design. In Downing, S. & Haladyna, T. (Eds) Handbook of test development. New York: Routledge

National Assessment Governing Board. (2008) Science Framework for the 2009 National Assessment of Educational Progress. Washington, DC: NABG, US Department of Education.

National Research Council (2000). How People Learn: Brain, Mind, Experience, and School. Washington, DC: National Academy Press.

Pellegrino, J., Chudowsky, N. and Glaser R. (2001) Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.

Schmidt, W., Wang, H., McKnight, Curtis (2005) Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. *Journal of Curriculum Studies.* 37 (5) pp 525-559.

Schneider, M. (2009) *Diplomas and Dropouts: Which Colleges Actually Graduate Their Students (and Which Don't).* Washington DC: American Enterprise Institute (AEI),

Shavelson, R.J., M.A. Ruiz-Primo, and E.W. Wiley (2005). "Windows Into the Mind." *Higher Education*, 49:413–430.

Shepard, L. A. (2005) Formative assessment: Caveat emptor. ETS Invitational Conference. New York City, NY.

Webb, N. L. (1999). *Research monograph No. 18. Alignment of science and mathematics standards and assessments in four states.* Washington, DC: Council of Chief State School Officers.

Wilson, M. R., & Berenthal, M. W. (2005). *Systems for state science assessment.* Washington, D.C.,. National Academies Press.

Figure 1: An example of an integrated claims and evidence statement[2]

Big Idea 1: The process of evolution drives the diversity and unity of life.

EU 1A: Change in the genetic makeup of a population over time is evolution.

L3 1A.3: Evolutionary change is driven by genetic drift and artificial selection.

Skill 6.4: The student can make claims and predictions about natural phenomena based on scientific theories and models.

The Claim: The student can make predictions about the effects of natural selection versus genetic drift on the evolution of both large and small populations of organisms.

The Evidence: The work will include a prediction of the effects of either natural selection or genetic drift on two populations of the same organism, but of different sizes; the prediction includes a description of the change in the gene pool of a population; the work shows correctness of connections made between the model and the prediction and the model and the phenomena (e.g. genetic drift may not happen in a large population of organisms; both natural selection and genetic drift result in the evolution of a population).

---

[2] Taken from Huff, 2009. CCSSO Assessment Meeting, Los Angeles

Figure 2

ECD Activities and artifacts create a transparent evidentiary argument (Huff & Steinberg, 2009)