Defining and Assessing Competence in Science: *Factors Enabling a Dramatic Improvement in Science Education*

Jim Pellegrino

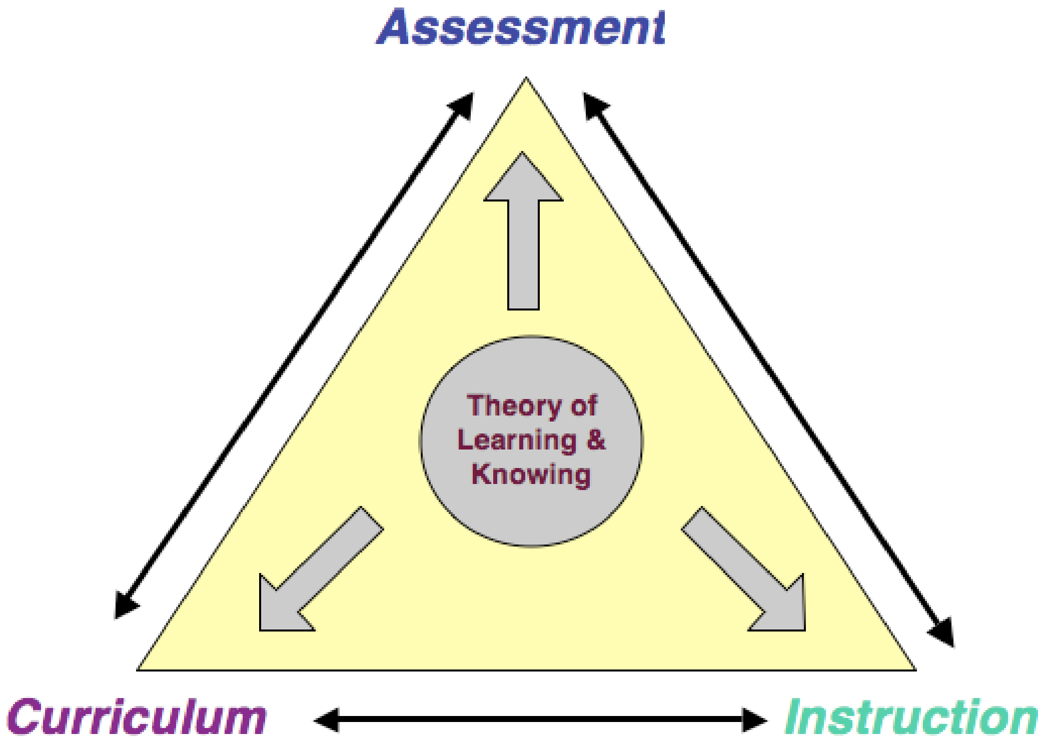
LEARNING SCIENCES

University of Illinois at Chicago
Learning Sciences Research Institute

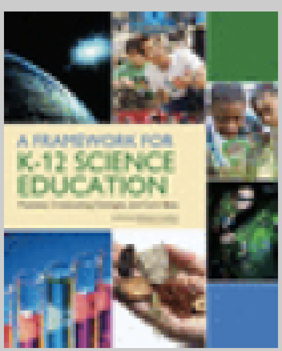
Main Points Today

- There is a real opportunity for a dramatic change in the nature of science assessment in the U.S. -- national, state, local, classroom
- The change is critical for U.S. science education
- Several factors seem to be coming together to herald and support such a change.
- At the heart of the change is a clearer description of what competence in science means and what it should look like in the classroom and on high-stakes tests.
- PER, NARST and other science education groups should be leading the way by defining opportunities and directions for R&D on assessment & learning.

Assessment Should not be the
"Tail that Wags the Educational Dog"



Source of My Optimism?



A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas

NRC Framework for Science Education Standards

- The Conceptual Framework for new science education standards has proposed a description of student competence as being the intersection of knowledge involving:
  - **scientific and engineering practices,**
  - **cross-cutting concepts**, and
  - **core disciplinary ideas,** with
  - **performance expectations** representing the intersection of core ideas and practices.

BOX ES.1
The Three Dimensions of the Framework

1. Scientific and Engineering Practices	**3. Disciplinary Core Ideas**
1. Asking questions (for science) and defining problems (for engineering) 2. Developing and using models 3. Planning and carrying out investigations 4. Analyzing and interpreting data 5. Using mathematics and computational thinking 6. Constructing explanations (for science) and designing solutions (for engineering) 7. Engaging in argument from evidence 8. Obtaining, evaluating, and communicating information **2. Crosscutting Concepts** 1. Patterns 2. Cause and effect: Mechanism and explanation 3. Scale, proportion, and quantity 4. Systems and system models 5. Energy and matter: Flows, cycles, and conservation 6. Structure and function 7. Stability and change	<u>Physical Sciences</u> PS 1: Matter and its interactions PS 2: Motion and stability: Forces and interactions PS 3: Energy PS 4: Waves and their applications in technologies for information transfer <u>Life Sciences</u> LS 1: From molecules to organisms: Structures and processes LS 2: Ecosystems: Interactions, energy, and dynamics LS 3: Heredity: Inheritance and variation of traits LS 4: Biological evolution: Unity and diversity <u>Earth and Space Sciences</u> ESS 1: Earth's place in the universe ESS 2: Earth's systems ESS 3: Earth and human activity <u>Engineering, Technology, and the Applications of Science</u> ETS 1: Engineering design ETS 2: Links among engineering, technology, science, and society

BOX 5-1
Core and Component Ideas in the Physical Sciences

Core Idea PS1: Matter and Its Interactions
- PS1.A: Structure and Properties of Matter
- PS1.B: Chemical Reactions
- PS1.C: Nuclear Processes

Core Idea PS2: Motion and Stability: Forces and Interactions
- PS2.A: Forces and Motion
- PS2.B: Types of Interactions
- PS2.C: Stability and Instability in Physical Systems

Core Idea PS3: Energy
- PS3.A: Definitions of Energy
- PS3.B: Conservation of Energy and Energy Transfer
- PS3.C: Relationship Between Energy and Forces
- PS3.D: Energy in Chemical Processes and Everyday Life

Core Idea PS4: Waves and Their Applications in Technologies for Information Transfer
- PS4.A: Wave Properties
- PS4.B: Electromagnetic Radiation
- PS4.C: Information Technologies and Instrumentation.

PS1.A: Structure and Properties of Matter				
	By the End of Grade 2	By the End of Grade 5	By the End of Grade 8	By the End of Grade 12
Tasks	Students support claims as to whether something is a solid or a liquid by providing descriptive evidence. Note: It is inappropriate at this grade level to use a substance, such as sand, that is made of visible scale particles but flows as the test material for this question. Test examples should be readily classifiable.	Students provide strategies for collecting evidence as to whether matter still exists when it is not visible.	Students create atomic and molecular models to explain the differences between the solid, liquid, and gaseous state of a substance.	Students first develop models that describe a neutral atom and a negative or positive ion. They then use these models to describe the similarities and differences between the atoms of neighboring elements in the periodic table (side by side or one above the other).
Criteria	Descriptive evidence that a material is a solid would include the object's definite shape; for a liquid it would be that the material takes the shape of the container or that the material flows to the lowest part of the container.	Design includes ways to measure weight with and without an invisible substance (gas or solute) present. For example, weighing the same container with different amounts of air, such as an inflated and deflated balloon or basketball; or weighing pure water and sugar before and after the sugar is dissolved in the water.	The model should show that atoms/molecules in a solid (1) are close together; (2) are limited in motion but vibrate in place; (3) cannot move past or around each other and thus are fixed in relative position. The model should also show that atoms/molecules in a liquid (1) are about as close together as in a solid, (2) are always disordered, (3) have greater freedom to move	The models should show that the atom consists of an inner core called the nucleus, which consists of protons and neutrons; that the number of protons in the nucleus is the atomic number and determines the element; that the nucleus is much smaller in size than the atom; that the outer part of the atom contains electrons; that in a neutral atom, the number of

Where do these ideas about the nature of competence and its development come from?

What does this work imply for the future and for individuals such as us?

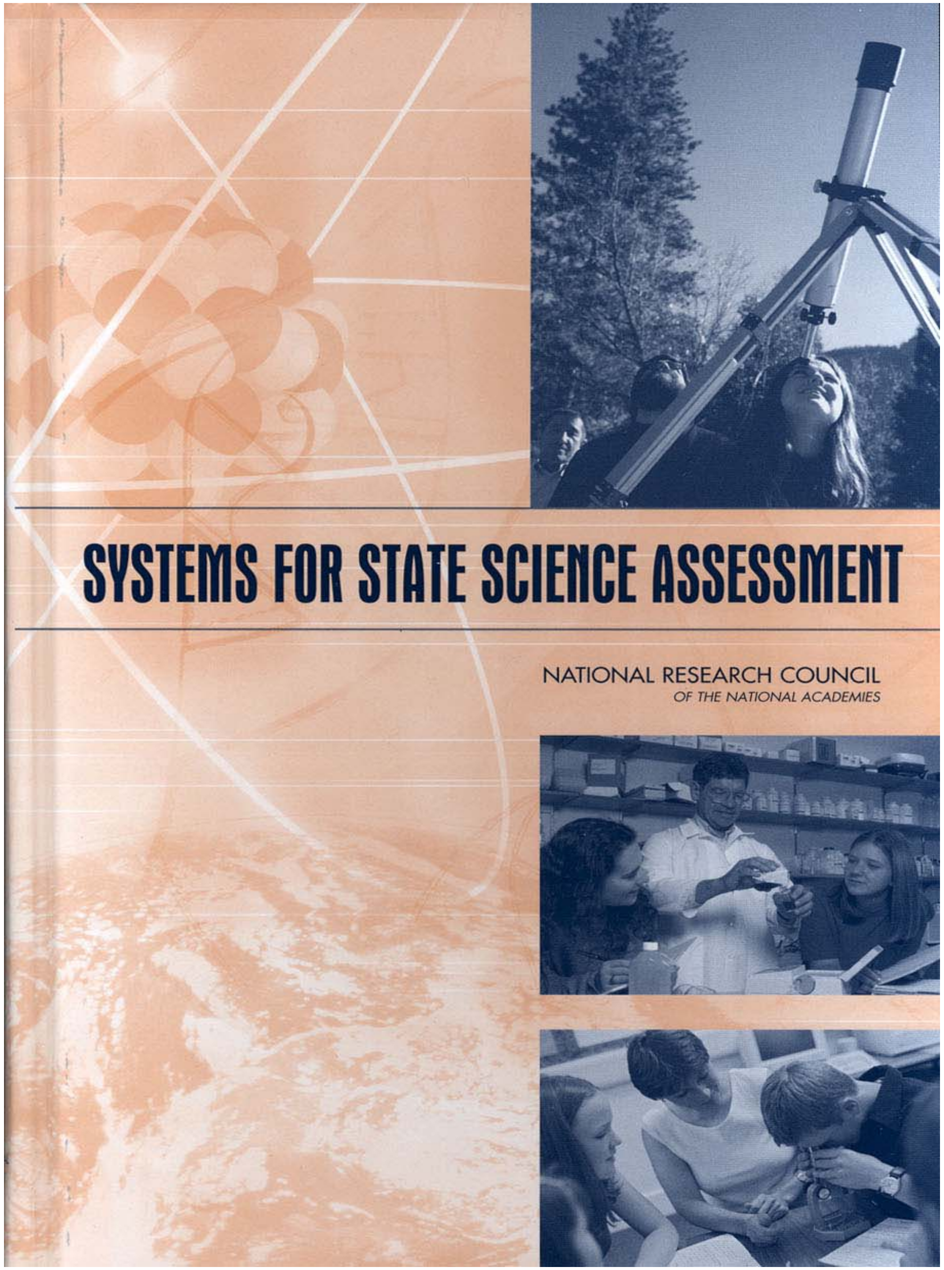
Two Major Sources of Influence

1. **Prior NRC & related Reports on Science Teaching, Learning & Assessment**
   - **Descriptions of the nature of competence and its development**
2. **Efforts by Other Stakeholders to Build Coherent Systems for Science Curriculum-Instruction-Assessment**
   - **Detailed specifications of the components of competence for purposes of designing curriculum, instruction and assessment**

1. Prior NRC Reports on Science Teaching, Learning & Assessment

*Systems for State Science Assessment

*Taking Science to School



- Desired end product is a multilevel system
  - Each level fullfills a clear set of functions and has a clear set of intended users of the assessment information
  - The assessment tools are designed to serve the intended purpose
    - Formative, summative or accountability
    - Design is optimized for function served
- The levels are articulated and conceptually coherent
  - They share the same underlying concept of what the targets of learning are at a given grade level and what the evidence of attainment should be.
  - They provide information at a “grain size” and on the “time scale” appropriate for translation into action.

What such a system might look like

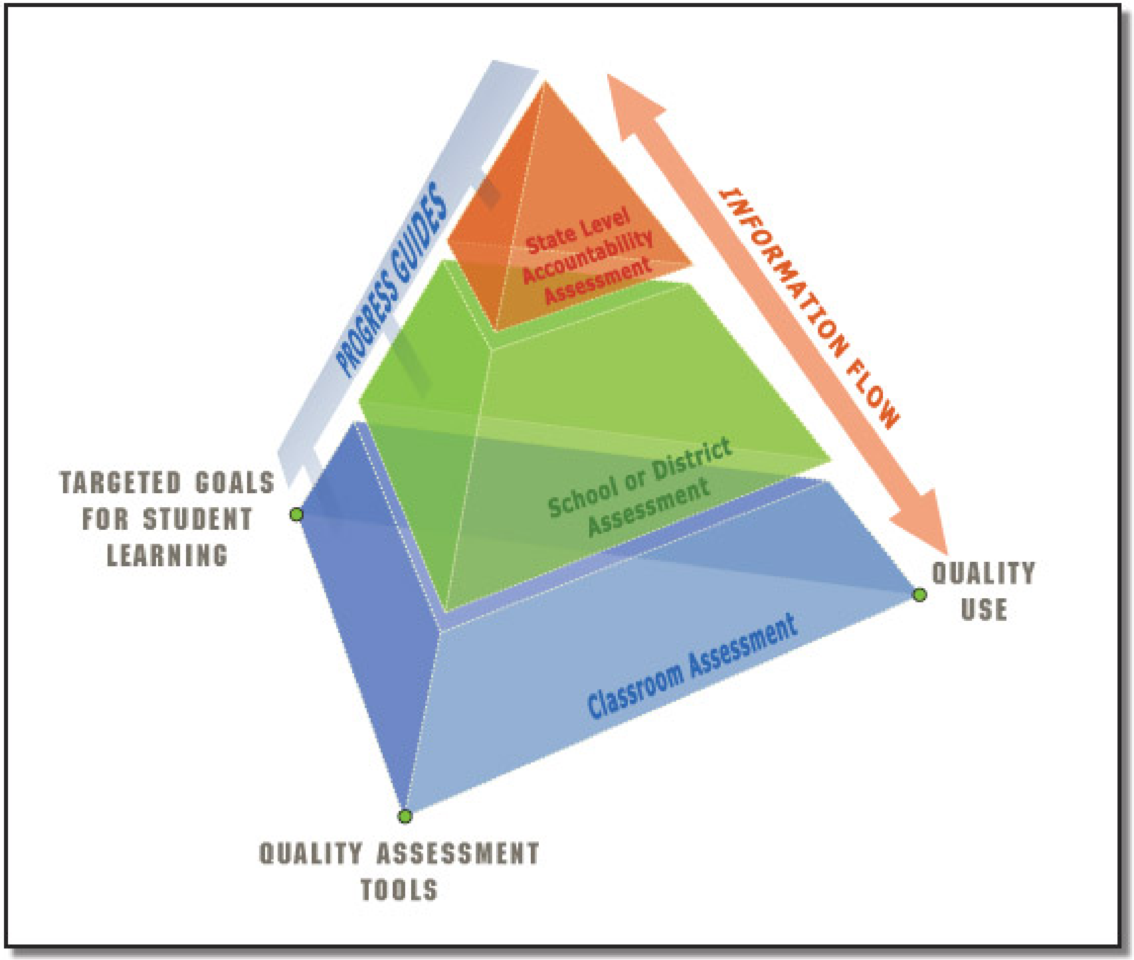
An Integrated System

Coordinated across levels

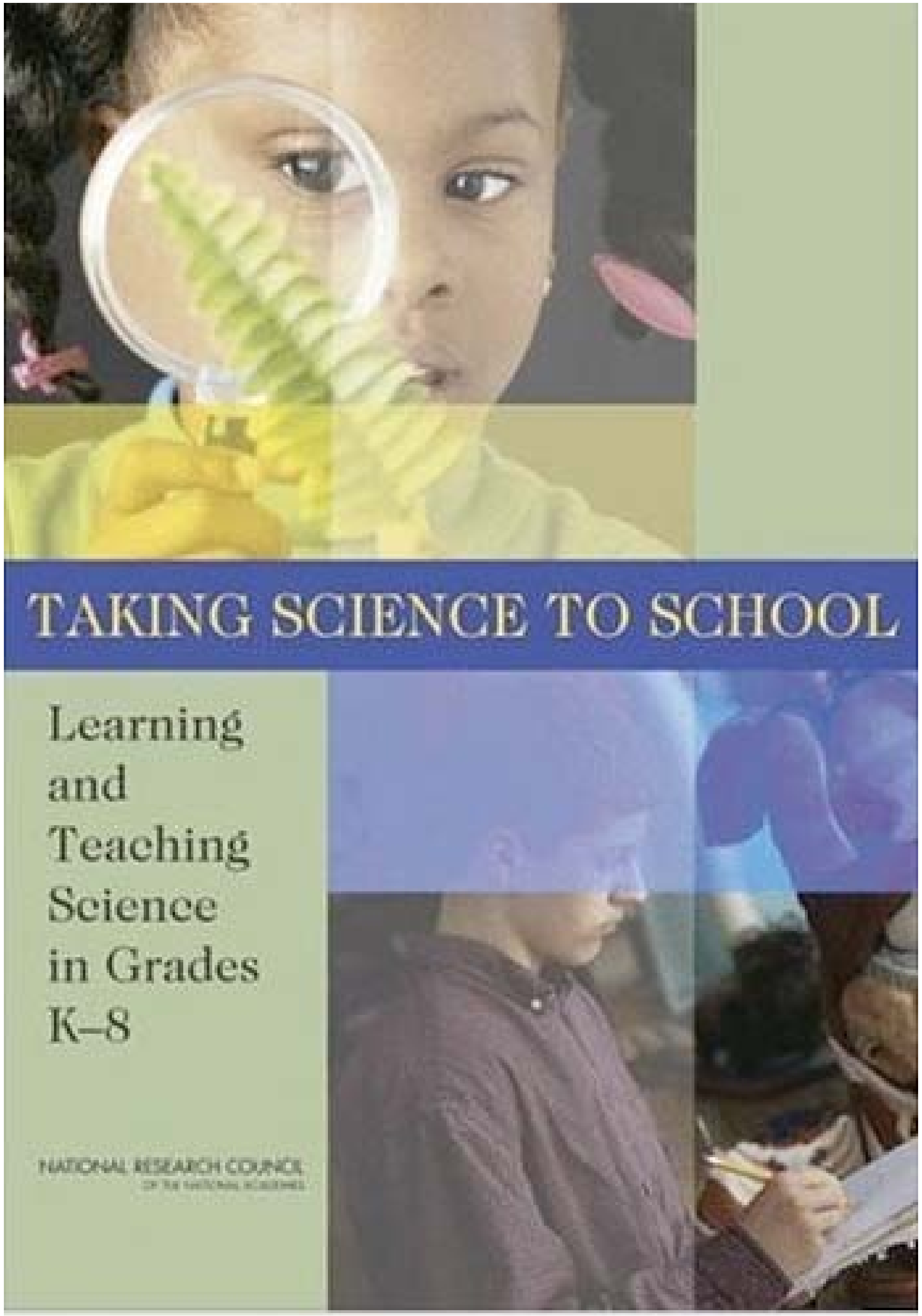
Unified by common learning goals

Synchronized by unifying progress variables

CAESL ASSESSMENT SYSTEM

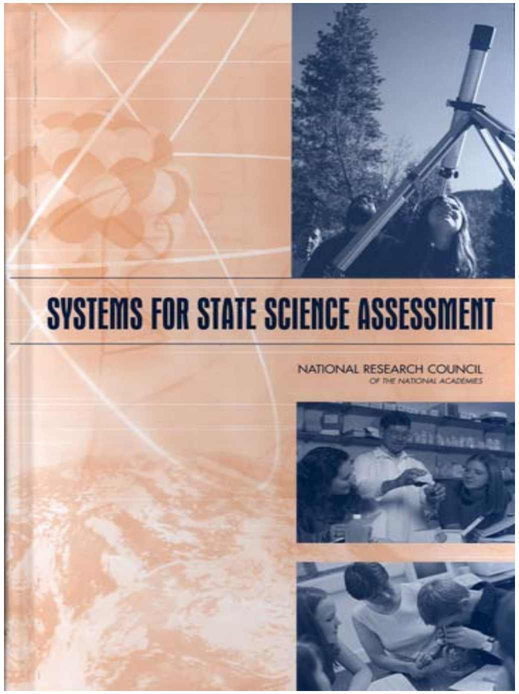


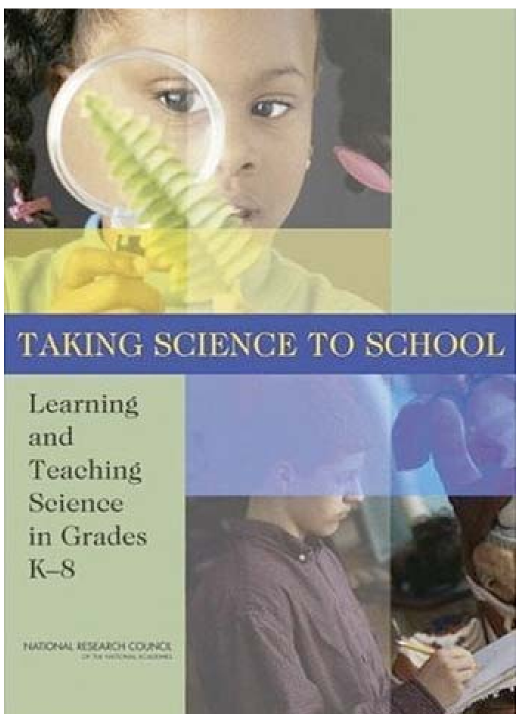
Multilevel Assessment System



Four Strands

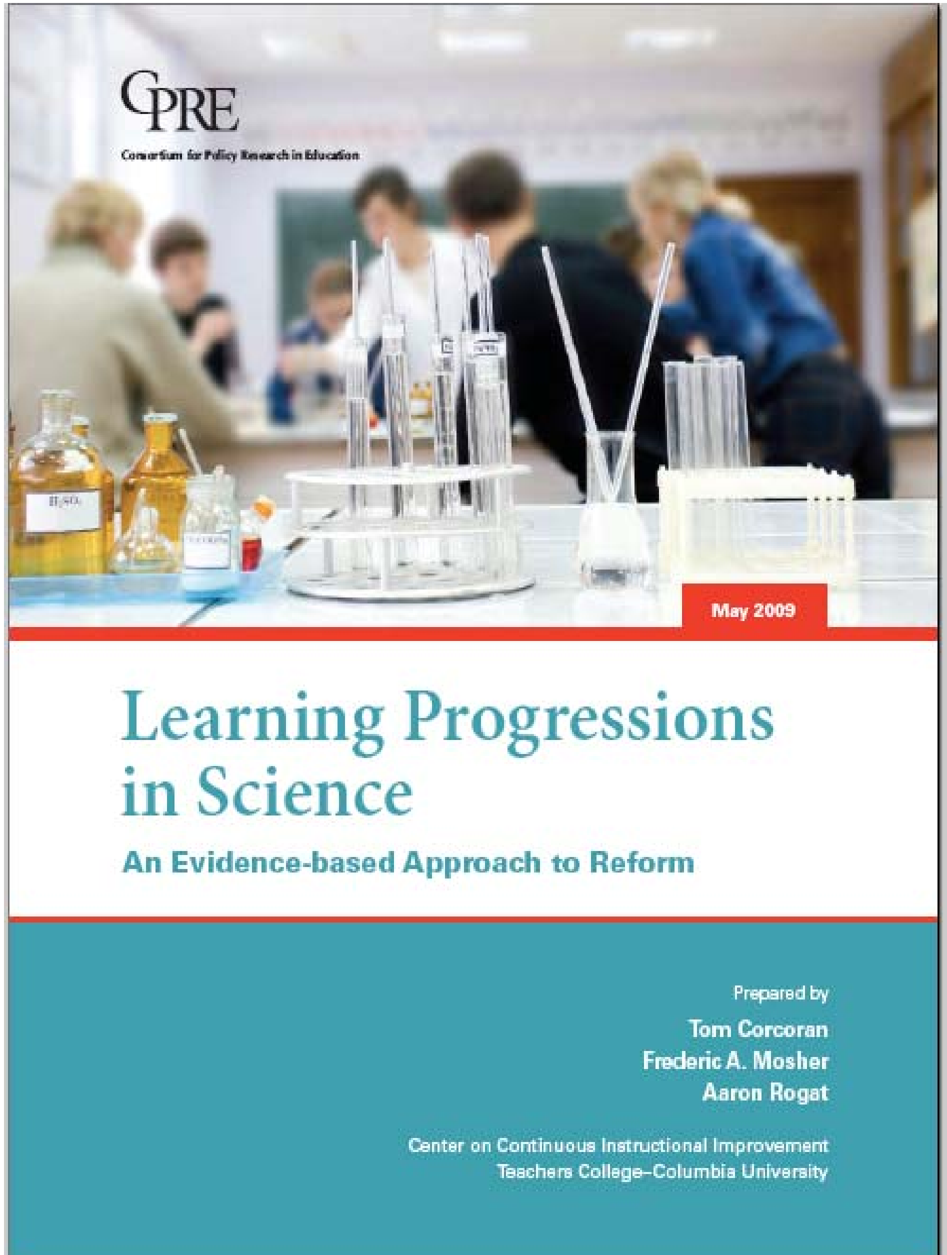
- **<u>Understanding scientific explanations</u>** with an emphasis on fundamental concepts rather than on memorizing unconnected facts. Students who are proficient in this strand know, use, and interpret scientific explanations of the natural world.
- **<u>Generating scientific evidence</u>**. This may include making observations, formulating a research question, developing a hypothesis (perhaps in the form of a model), using a range of methods to gather data, analysis of data and confirmation and/or revision of the hypothesis.
- **<u>Reflecting on science,</u>** which includes not only understanding the nature and development of scientific knowledge, but also reflecting on one's own learning and understanding of science.
- **<u>Participating productively in scientific practices and discourse</u>**. This strand flows out of the notion that science takes place within a community of practice that shares norms, practices, and a common language, and that learners should be introduced to these norms and practices as they experience and engage with science.





Learning Progressions: Descriptions of successively more sophisticated ways of thinking about key disciplinary concepts and practices across multiple grades

- Structured around big ideas and practices- powerful and generative
- Upper anchor- societal expectations of what students should know; based on analysis of discipline
- Lower anchor - what students come in with
- Describes how learning develops- the intermediate steps towards expertise
- Grounded in synthesis of education research



Potential Value of Learning Progressions

- LPs can guide the design of instruction
- LPs can guide the specification of learning performances - connecting disciplinary practices and "big ideas"
- Learning performances can guide the development of tasks that allow us to observe and infer students' level of competence for major constructs that are the target of instruction and assessment

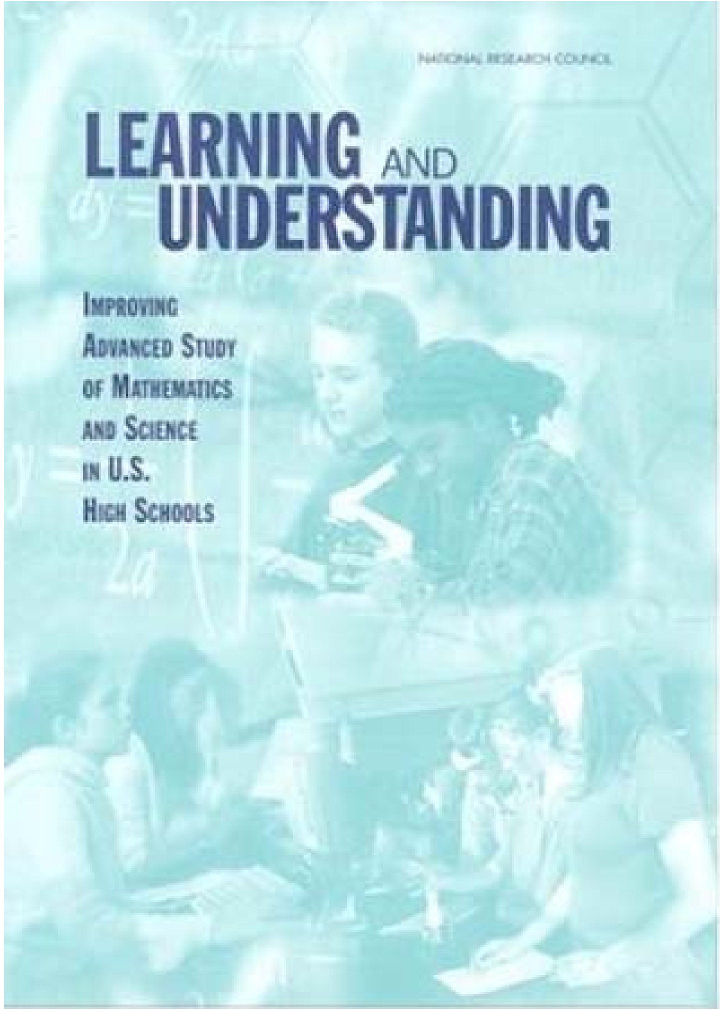
2. Efforts by Other Stakeholders to Build Coherent Systems for Science Curriculum-Instruction-Assessment

* NSF & College Board AP Science Redesign Project

* College Board Standards for College Success: Science

Why an AP Science Redesign?

- **A 2002 NRC Report identified ways to improve advanced study of math and science in the U.S. The Report's recommendations are applicable to all AP course subjects:**
  - Emphasize deep understanding rather than comprehensive coverage -- avoid "mile wide & inch deep" syndrome
  - Reflect current understanding of how students learn in a discipline
  - Reflect current research directions within the disciplines
  - Emphasize the development of inquiry and reasoning skills



Conceptual Approach of the Redesign Built Upon a Variety of Work:

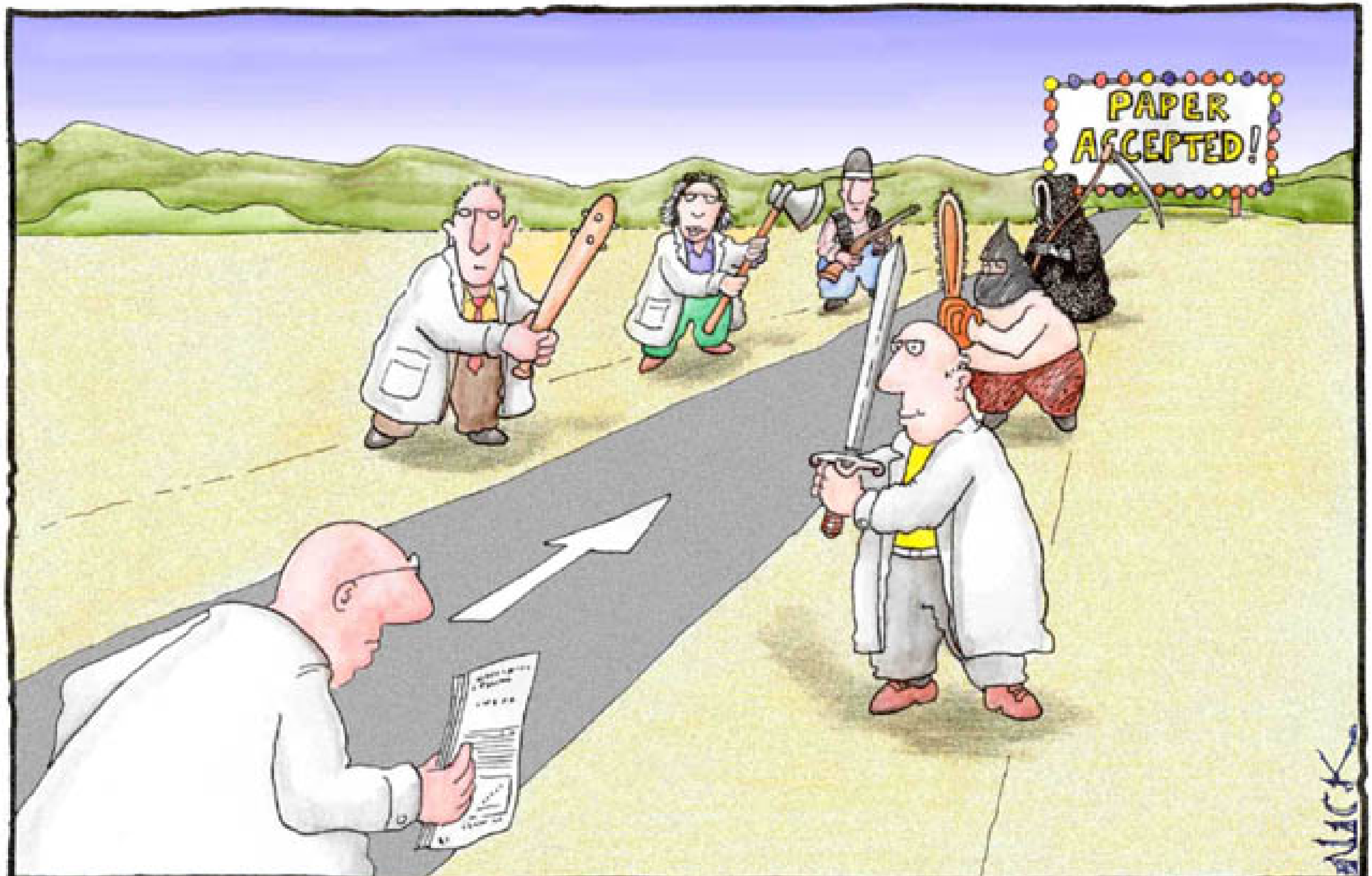
Understanding by Design, G. Wiggins & J. McTighe (1998, 2006).

How People Learn: Brain, Mind, Experience & School. J. Bransford, A. Brown, & R. Cocking, S. Donovan, & J. Pellegrino (2000).

Knowing What Students Know: The Science and Design of Educational Assessment, J. Pellegrino, N. Chudowsky, & R. Glaser (2001).

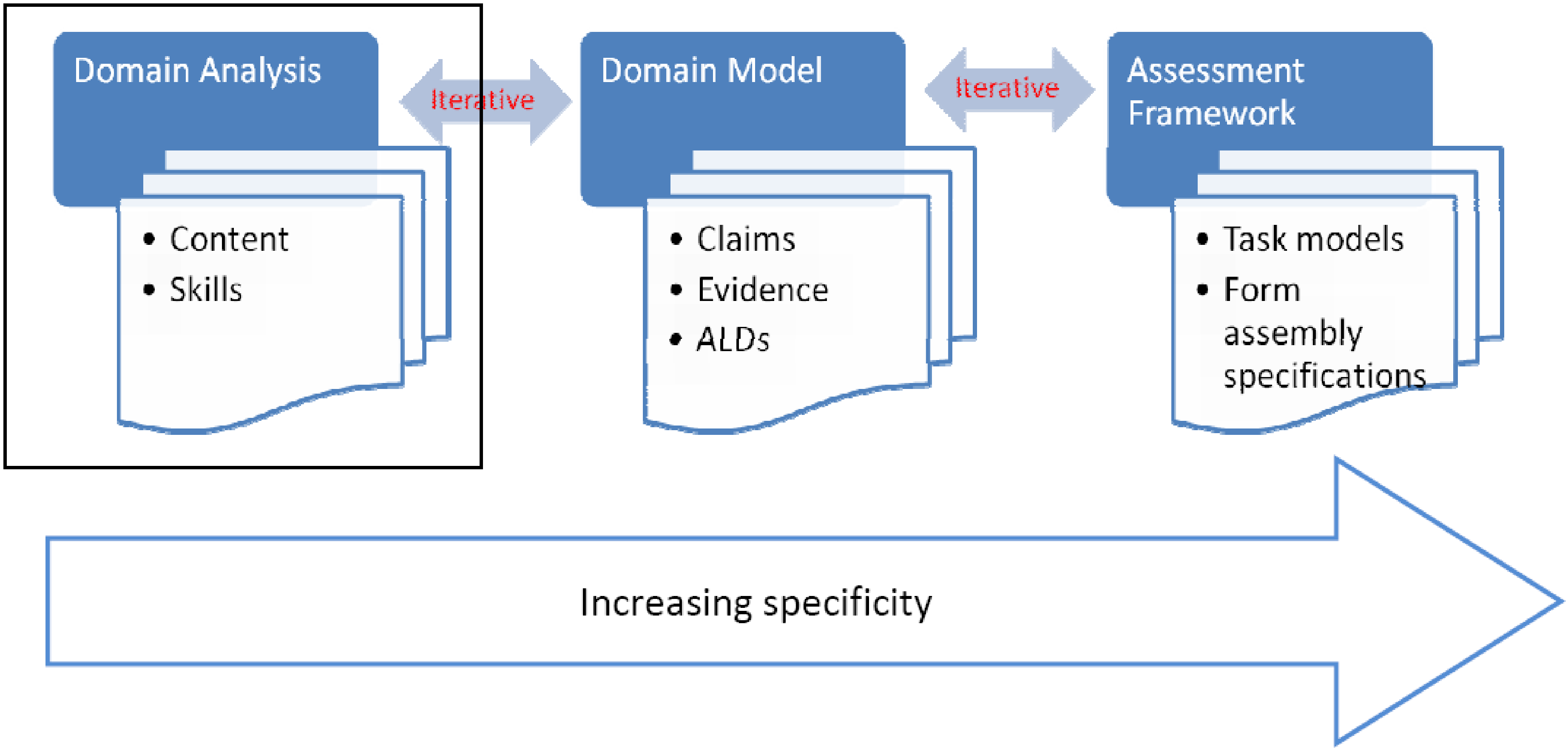
Evidence-Centered Assessment Design: Layers, Structures, and Terminology, R.J. Mislevy & M.M. Riconscente (2005).

Development & Peer Review by Domain Experts

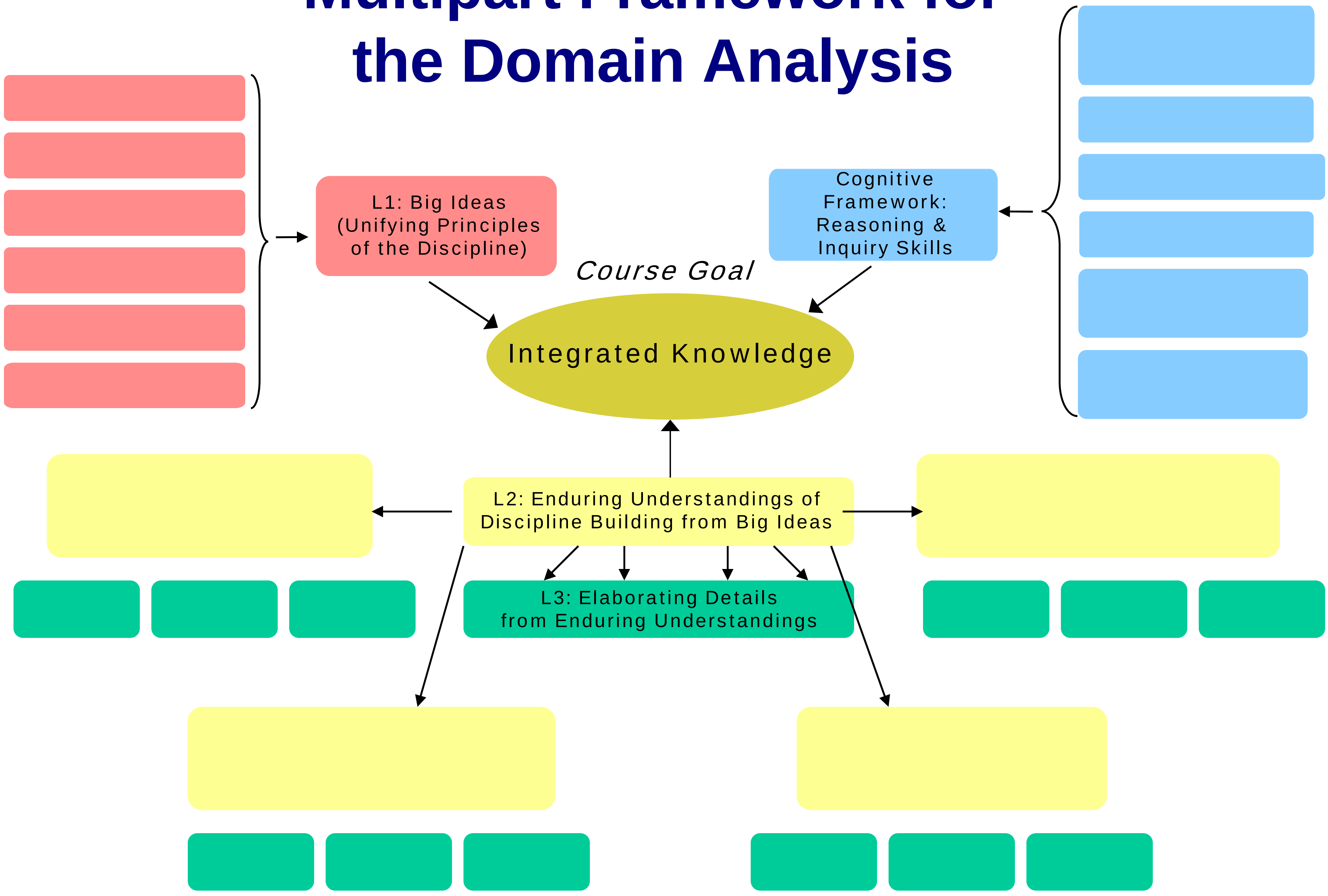


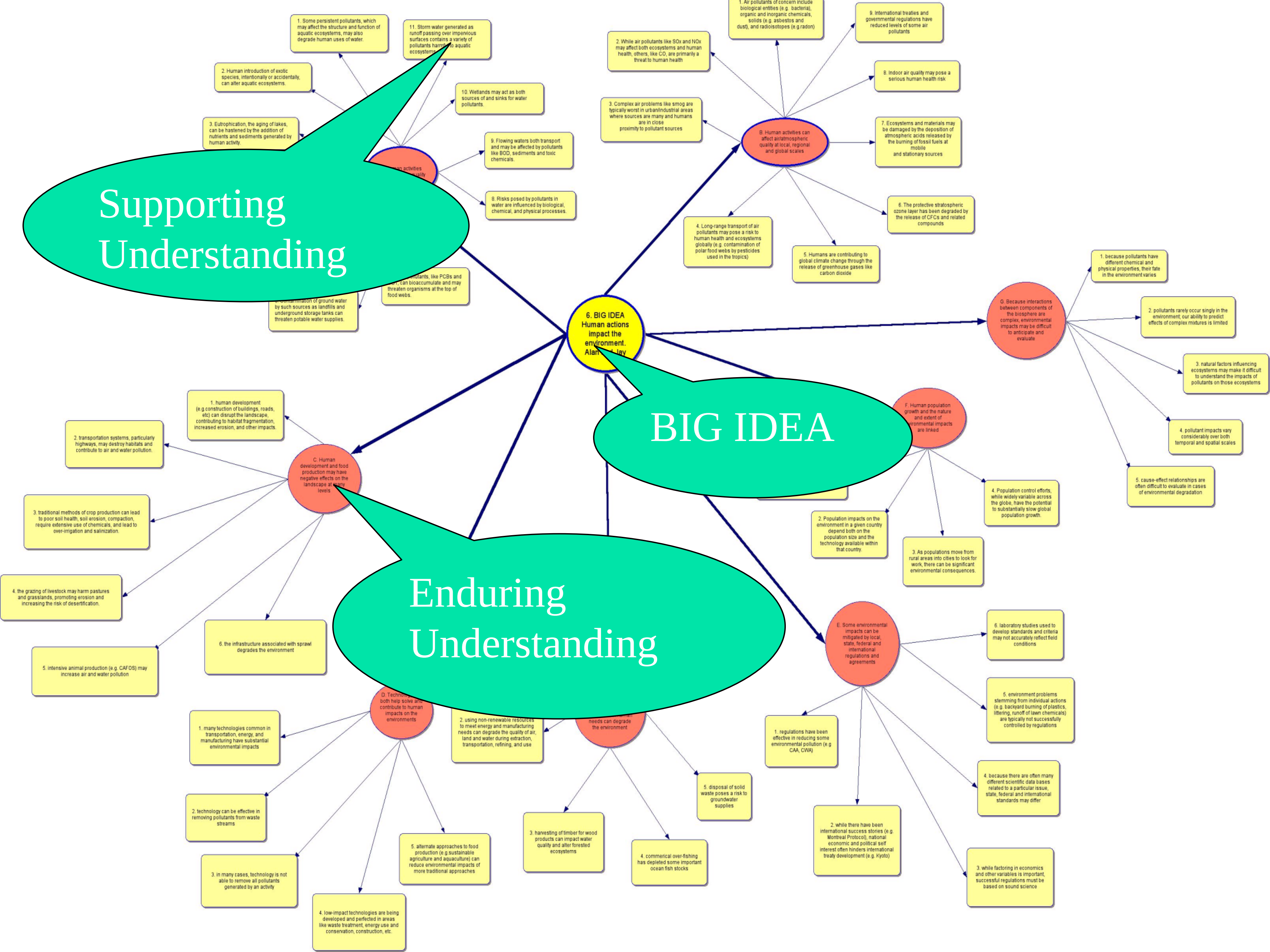
Most scientists regarded the new streamlined peer-review process as 'quite an improvement.'

Three Critical Design Phases



Multipart Framework for the Domain Analysis




Supporting Understanding
BIG IDEA
Enduring Understanding
6. BIG IDEA Human actions impact the environment.
B. Human activities can affect air/atmospheric quality at local, regional and global scales
1. Air pollutants of concern include biological entities (e.g. bacteria), organic and inorganic chemicals, solids (e.g. asbestos and dust), and radioisotopes (e.g.radon)
2. While air pollutants like SOx and NOx may affect both ecosystems and human health, others, like CO, are primarily a threat to human health
3. Complex air problems like smog are typically worst in urban/industrial areas where sources are many and humans are in close proximity to pollutant sources
4. Long-range transport of air pollutants may pose a risk to human health and ecosystems globally (e.g. contamination of polar food webs by pesticides used in the tropics)
5. Humans are contributing to global climate change through the release of greenhouse gases like carbon dioxide
6. The protective stratospheric ozone layer has been degraded by the release of CFCs and related compounds
7. Ecosystems and materials may be damaged by the deposition of atmospheric acids released by the burning of fossil fuels at mobile and stationary sources
8. Indoor air quality may pose a serious human health risk
9. International treaties and governmental regulations have reduced levels of some air pollutants
1. Some persistent pollutants, which may affect the structure and function of aquatic ecosystems, may also degrade human uses of water.
2. Human introduction of exotic species, intentionally or accidentally, can alter aquatic ecosystems.
3. Eutrophication, the aging of lakes, can be hastened by the addition of nutrients and sediments generated by human activity.
8. Risks posed by pollutants in water are influenced by biological, chemical, and physical processes.
9. Flowing waters both transport and may be affected by pollutants like BOD, sediments and toxic chemicals.
10. Wetlands may act as both sources of and sinks for water pollutants.
G. Because interactions between components of the biosphere are complex, environmental impacts may be difficult to anticipate and evaluate
1. because pollutants have different chemical and physical properties, their fate in the environment varies
2. pollutants rarely occur singly in the environment, our ability to predict effects of complex mixtures is limited
3. natural factors influencing ecosystems may make it difficult to understand the impacts of pollutants on those ecosystems
4. pollutant impacts vary considerably over both temporal and spatial scales
5. cause-effect relationships are often difficult to evaluate in cases of environmental degradation
2. Population impacts on the environment in a given country depend both on the population size and the technology available within that country.
3. As populations move from rural areas into cities to look for work, there can be significant environmental consequences.
4. Population control efforts, while widely variable across the globe, have the potential to substantially slow global population growth.
C. Human development and food production may have negative effects on the landscape at many levels
1. human development (e.g.construction of buildings, roads, etc) can disrupt the landscape, contributing to habitat fragmentation, increased erosion, and other impacts.
2. transportation systems, particularly highways, may destroy habitats and contribute to air and water pollution.
3. traditional methods of crop production can lead to poor soil health, soil erosion, compaction, require extensive use of chemicals, and lead to over-irrigation and salinization.
4. the grazing of livestock may harm pastures and grasslands, promoting erosion and increasing the risk of desertification.
5. intensive animal production (e.g. CAFOS) may increase air and water pollution
6. the infrastructure associated with sprawl degrades the environment
E. Some environmental impacts can be mitigated by local, state, federal and international regulations and agreements
1. regulations have been effective in reducing some environmental pollution (e.g. CAA, CWA)
2. while there have been international success stories (e.g. Montreal Protocol), national economic and political self interest often hinders international treaty development (e.g. Kyoto)
3. while factoring in economics and other variables is important, successful regulations must be based on sound science
4. because there are often many different scientific data bases related to a particular issue, state, federal and international standards may differ
5. environment problems stemming from individual actions (e.g. backyard burning of plastics, littering, runoff of lawn chemicals) are typically not successfully controlled by regulations
6. laboratory studies used to develop standards and criteria may not accurately reflect field conditions
1. many technologies common in transportation, energy, and manufacturing have substantial environmental impacts
2. technology can be effective in removing pollutants from waste streams
3. in many cases, technology is not able to remove all pollutants generated by an activity
4. low-impact technologies are being developed and perfected in areas like waste treatment, energy use and conservation, construction, etc.
5. alternate approaches to food production (e.g.sustainable agriculture and aquaculture) can reduce environmental impacts of more traditional approaches
2. using non-renewable resources to meet energy and manufacturing needs can degrade the quality of air, land and water during extraction, transportation, refining, and use
3. harvesting of timber for wood products can impact water quality and alter forested ecosystems
4. commercial over-fishing has depleted some important ocean fish stocks
5. disposal of solid waste poses a risk to groundwater supplies

AP Science Reasoning: Level 1

1. Use representations and models to communicate scientific phenomena and solve scientific problems.
2. Use mathematics appropriately.
3. Engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.
4. Plan and implement data collection strategies in relation to a particular scientific question.
5. Perform data analysis and evaluation of evidence.
6. Work with scientific explanations and theories.
7. Connect and relate knowledge across various scales, concepts, and representations in and across domains.

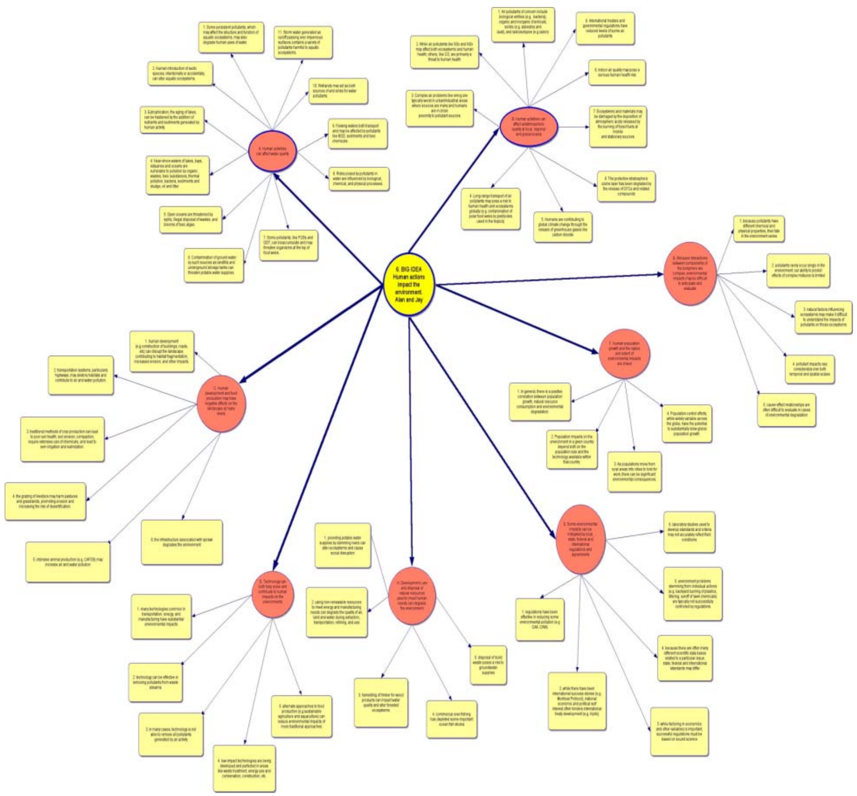
AP Science Reasoning: Level 2

Level 1: work with scientific explanations & theories

Level 2:

1. justify claims with evidence
2. construct explanations of phenomena based on evidence produced through scientific practices
3. articulate the reasons that scientific explanations and theories are refined or replaced
4. make claims and predictions about natural phenomena based on scientific theories & models.
5. evaluate alternative scientific explanations

Content & Science Reasoning



1.3 The student can re-express key elements of natural phenomena across multiple representations in the domain.

1.3.1 Correspondence of elements of one representation with another representation

1.3.2 Accuracy of description

1.3.3 Correctness of focus on specific essential qualities

1.3.4 Differentiation from similar terms/phenomena

1.3.5 Correspondence between an explanation and a visual representation that is produced by instruments

1.3.6 Appropriateness of a visual representation used to explain a familiar phenomenon

1.3.7 Appropriateness of a visual representation used to explain a novel phenomenon in a familiar domain

1.3.8 Correctness of the scientific language used in a verbal explanation or narrative, and correspondence at the appropriate level of detail of the features of phenomena that are presented as visual representations (schemata, diagrams, graphs) that are precise and contains (identifies) significant features.

1.3.9 Correctness and precision of visual symbolic representation that is representative of language in a word problem (e.g., a vector diagrams of the forces involved a mechanics problems; a chemical reaction as a change in molecules, drawn as groups of atoms with their orbits) is included.

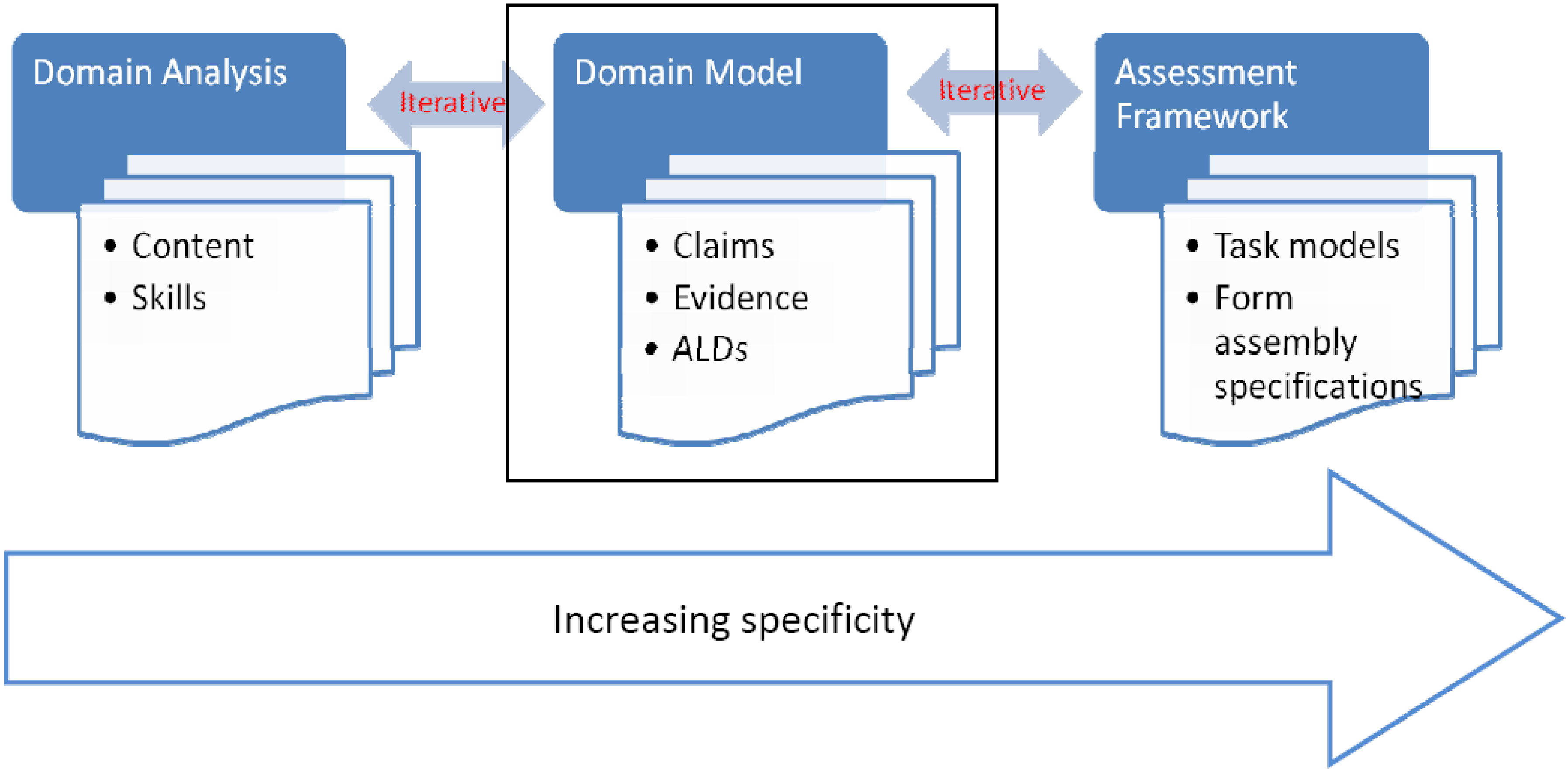
1.3.10 Appropriateness of a basic causal explanation expressed as a visual conceptual model (e.g., the causal explanation of plate tectonics; a causal explanation of thermal dilation using the atomic -molecular model)

1.3.11 Correspondence between a qualitative and quantitative solution to a problem, using symbolic representations, (e.g., light rays propagating through different media, or different optical lenses)

Content

Science Reasoning

Three Critical Design Phases



Evidence-Centered Design

Exactly what knowledge do you want students to have and ***how*** do you want them to know it?

What will you accept as ***evidence*** that a student has the desired knowledge?

How will you analyze and interpret the evidence?

What task(s) will the students ***perform*** to communicate their knowledge?

claim space

evidence

task

Instantiated through the Intersection of Course Content & Science Reasoning

Illustrative Claims and Evidence
AP Chemistry

Big Idea 2: Chemical & physical properties can be explained by the structure/arrangement of atoms, ions or molecules & the forces btw them.

EU 2C: The strong electrostatic forces of attraction holding atoms together in a unit are called chemical bonds.

L3 2C.1: In covalent bonding, electrons are shared between the nuclei of two atoms to forma a molecule or polyatomic ion.

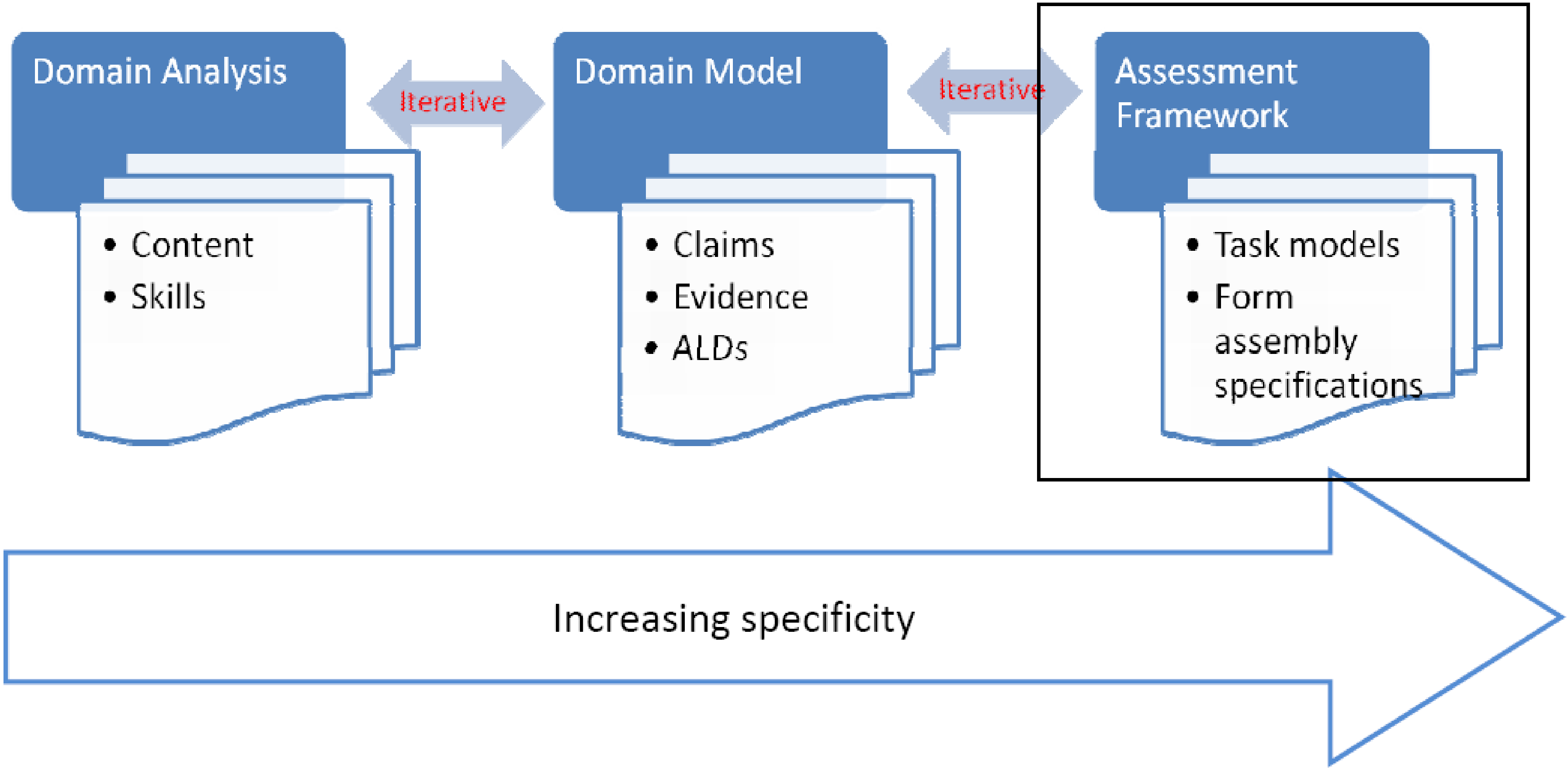
Skill 6.1: The student can justify claims with evidence.

The Claim: The student is able to use evidence to justify claims involving the classification of covalent bonds in terms of polarity and the relative strength of covalent bonds.

The Evidence: Justification accurately points to difference in values of electronegativity as evidence supporting classification as polar or nonpolar. (6.1.1) Justification includes link between electronegativity and distribution of shared electrons within a molecule. (6.1.2) Appropriateness of reasoning that electronegativity values for the representative elements increase going from left to right across a period and decrease going down a group. (6.1.2) Relative bond strength claims supported by values of bond energy and/or electronegativity and/or numbers of shared electrons and/or graphical representations of distance vs. potential energy. (6.1.1) Relative bond length claims are supported by values of bond length vs. numbers of electrons shared. (6.1.1) Interpretation of graphical representations of distance vs. potential energy consistent with Coulomb's law. (6.1.1) Inclusion and reasonableness of a statement that two or more valence electrons shared between atoms of unequal electronegativity constitute a polar covalent bond. (6.1.7)

Achievement Level: 3

Three Critical Design Phases



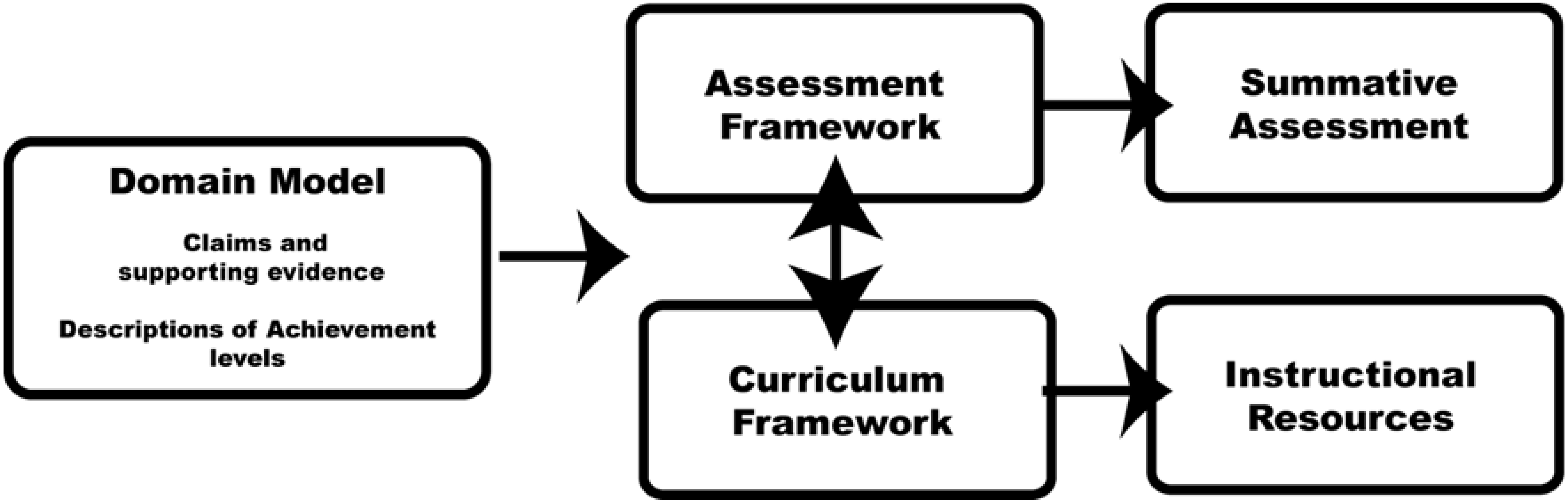
Task Models:
The Basis for Item Design

Provide the explicit link between the claims and evidence and the items.

- Directions
- Stimulus Material
- Prompt
- Response Options (MC)
- Student work product (CR)

Support validity of score inferences

Connecting the Domain Model to Curriculum, Instruction, & Assessment



Lessons Learned

- ***No Pain -- No Gain!!!*** -- this is hard work
- ***Backwards Design*** and ***Evidence Centered Design*** are challenging to execute & sustain
  - Require multidisciplinary teams
  - Require sustained effort and negotiation
  - Require time and money
- ***Value-added*** -- Validity is “designed in” from the start as opposed to “grafted on”
  - Elements of a validity argument are contained in the process and the products

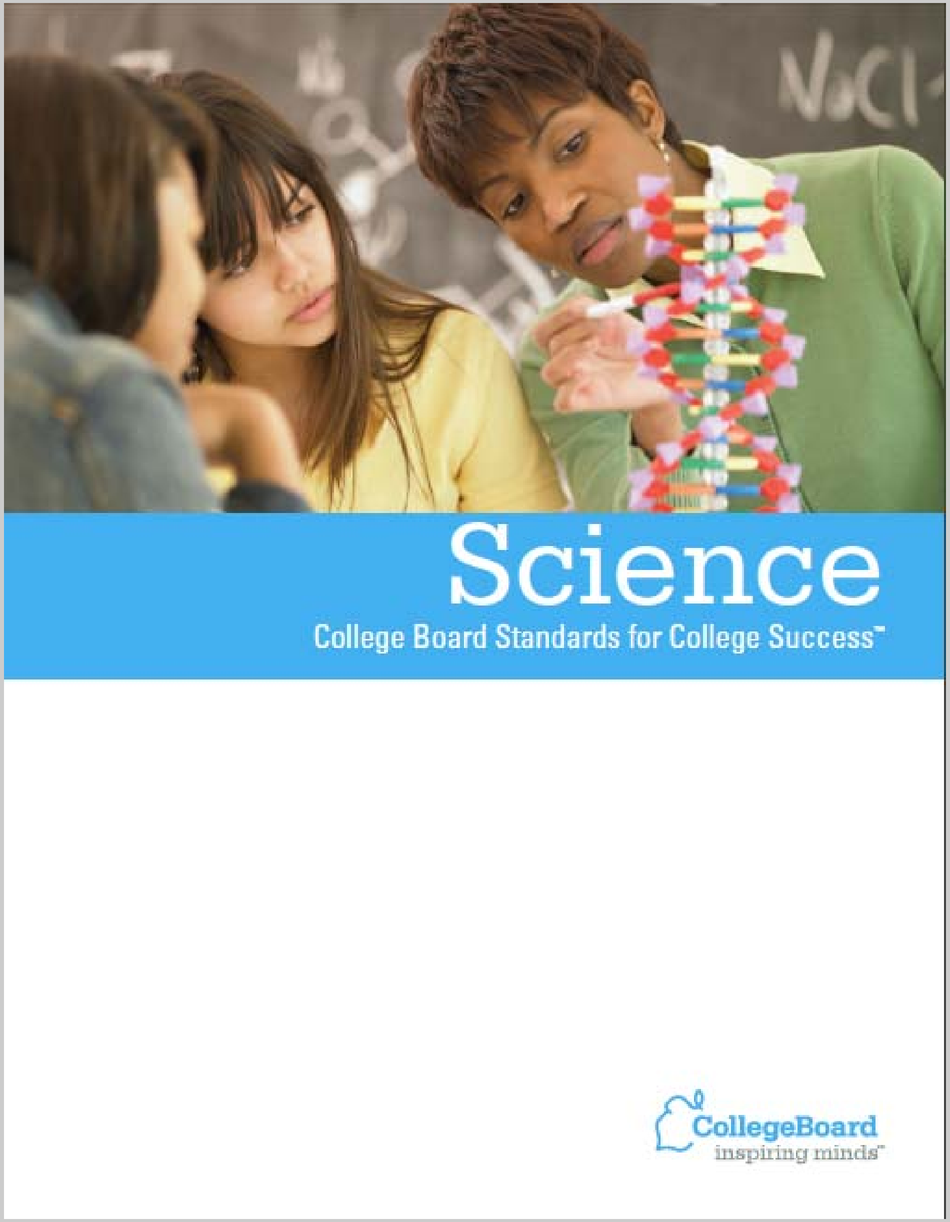
AP Redesign Implications

For AP science teachers and students:

- AP instructors and students will have a well-defined set of learning objectives that support teaching for deeper understanding.
- The AP Exams will be congruent with these learning objectives.
- AP instructors will have tools and professional development opportunities that support teaching, learning and success on the AP Exam
- The post-secondary community and professional societies will have a better understanding of, and confidence in, the value of AP courses.

For science education generally:

- The emphasis on reasoning and inquiry, enforced by the exams, can influence school science in lower grades and in the universities.
- The development of a high-stakes exam based on evidence-centered design principles can influence state and national assessment designs.
- The redesigned courses will increase interest and success within a new population of students who can then contribute to both science education and the practice of science.



- Standards for middle school through high school
- Builds from and toward the science foci in the AP course frameworks
- Uses the “Big Ideas” and “Science Practices” of the AP frameworks
- Emphasizes the description of performance expectations that integrate content and reasoning

Where Do We Stand?

- Multiple sources of wisdom and hard work have brought us to the present point of opportunity: the NRC Framework
- We have a much better sense of what the development of competence should mean and the possible implications for designing coherent science education
- We have examples of thinking through in detail what it means to juxtapose ***science practices*** and ***core content knowledge*** to guide the design of assessment

What's Left to Do?

- We have a lot of work to do to make sure that the next generation science standards live up to expectations
- We need to work together to translate the standards into effective models, methods and materials for curriculum, instruction, and assessment.
- We need to use what we know already to **evaluate** and **improve** the assessments that are part of current practice, e.g., concept inventories, large-scale exams, etc.

Final Reminder:
Assessment Should not be the
"*Tail that Wags the Educational Dog*"