Assessing Complex Human Performance

Eva L. Baker, Co-Director

UCLA CRESST

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

Mathematical Sciences Research Institute – Berkeley, California

April 2013

The Story

With lots of trials and many errors in the past, the intrepid assessor (with her monkey) wants to be clear about what she is or is not creating, how learners' thinking ,performance may (or may not) vary, and how to create the evidence needed to trust results

To do so, she must make rational choices and find tools to help her. Enter her mind!

Complexity in Assessment

- Purposes
- Beliefs
- Formats
- Features of complexity
- Tools
- Quality

Remember

- To use this opportunity to fix high stakes testing and teaching
- No presentation content suggests that any aspects should be assessed separately
- Intention is to add to or confirm options you have
- Tasks must be models that allow multiple similar instances
- Tasks must be transparent in design to influence teaching and learning
- To be open to new ideas about psychometriccs, that is not what you have experienced already
- Or not

Purposes of Assessments in Learning

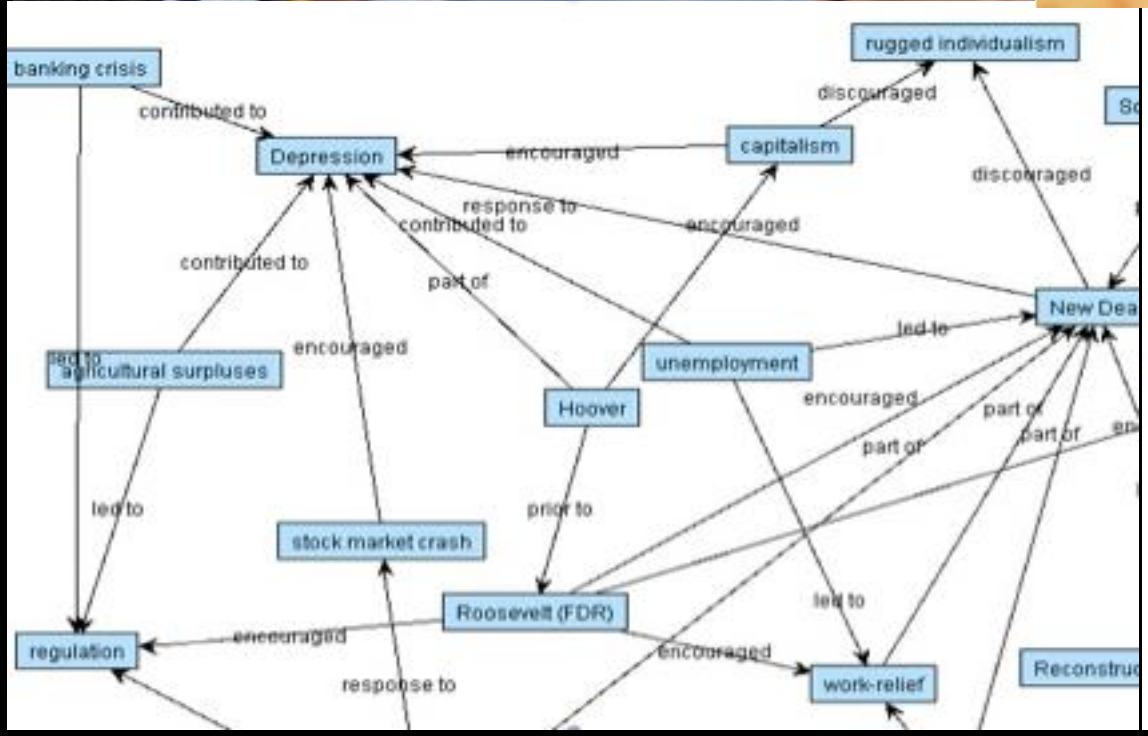
- Personalized or conforming
- Formative
  - Appropriate performance chunks
  - Inherent or supplementary support
  - Exploratory environments
  - Teacher/instruction interactivity has evidence-base
- Student feedback
  - Domain
  - Cognitive demands
  - Peers, self, exterenal resources, plus
  - Affective, efficacy, engagement

Purposes of Assessments in Learning

- Summative
  - Evidence of validity for purposes (accountability, certification, improvement)
  - Evidence of vertical scalingor equivalent
  - Evidence of classification reliability
  - Evidence of negative impact
  - Sensible summaries to support uses,instructional improvement, policy

My Choices: Beliefs

- Think about domain content only
- Assessment design is mostly art
- Formats are key
- Rules don't work
- Can't mix models
- ***High quality can be designed into assessments***
- ***All performance is idiosyncratic, but so what***
- ***Assess for what can be learned/taught***



Myth: Assessment is first & best defined by format

Choices: Types of Learner Performance

- Product, performance, combination
- Ephemeral, recorded, reported
- Self-contained, evolving
- Independent, with help, modeled
- Collaborative or team task
- Extended, interacting,
- Generated by learner with minimal prompts
- Uniform
- Consider complexity in tasks, scoring, instruction

What Are Common Thoughts on Complexity in Assessment Task Demands?

1. Difficulty *
2. Number of steps
3. Integration across principles, schema, and content
4. Close discriminations, e.g., confusability
5. Frequency of inferencing
6. Degree or onceptual distance of inferencing
7. Barriers to overcome, e.g., construct irrelevance
8. Features of situation, e.g., applications, fairness

How Can Complexity Play Out in Scoring?

- Alternative processes allowed
- Alternative processes differentially valued
- Differing acceptable outcomes
- Outcomes embedded in levels of content
- Sophistication in qualifications and evidence of scorers
  - Views of content, cognition and other attributes of performance
- Training, reliability, validity

How Is Complexity Related to Instruction?

- Prior knowledge requirements
- Potential to acquire key knowledge in assessment setting
- Instructional sensitivity (change attributed to explicit learning)
- Exposure continuum
  - Direct
  - Partial
  - Self-directed, given or found resources
- Communicatied for instrucional use

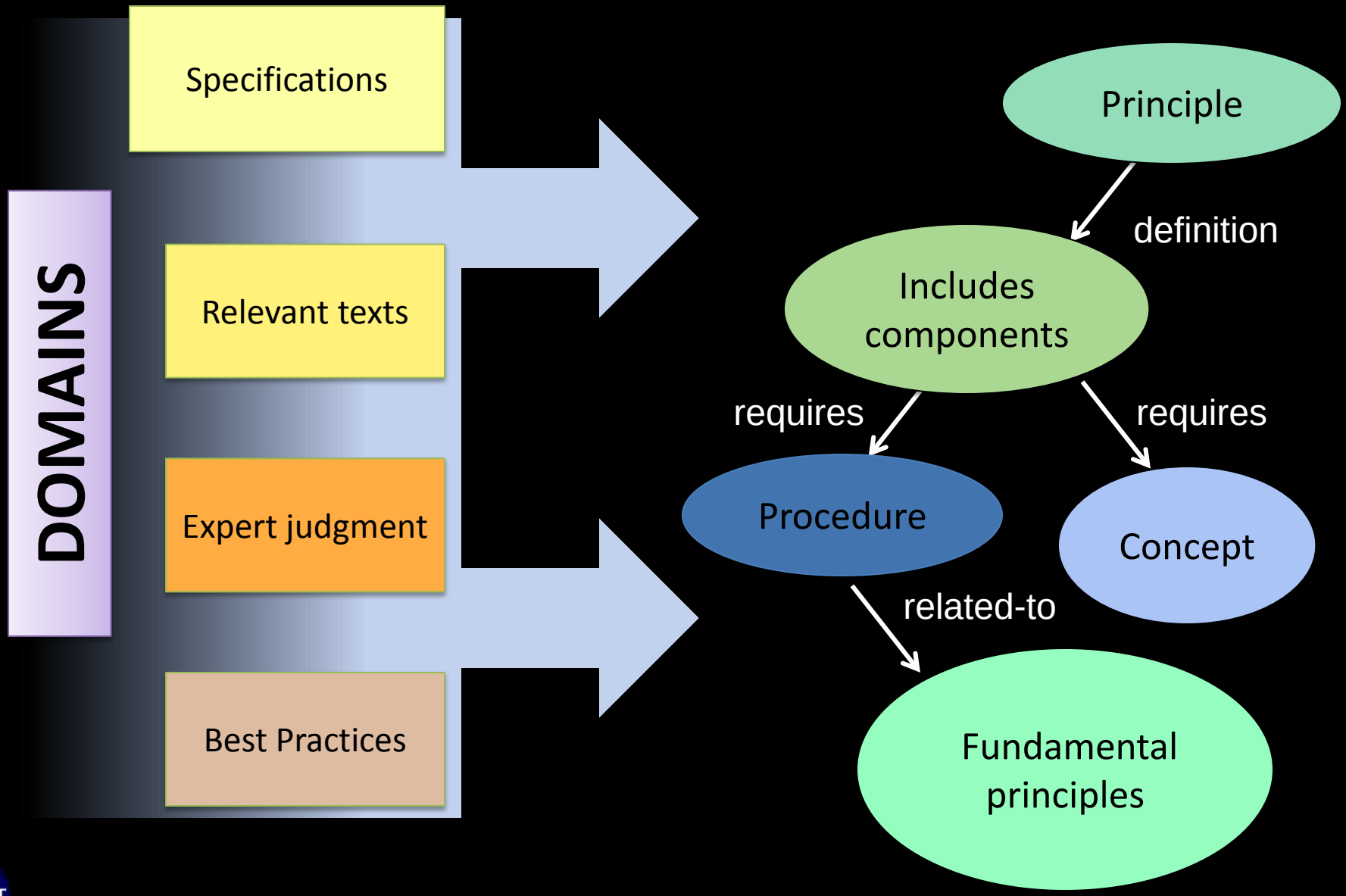
Complexity Displayed in Outcomes

- How features are selected and integrated?
- How they are sampled? Transparency?
- Comprehensiveness in view of standard
- Situations
- Practical utility
- Goals
  - Outcomes of instruction
  - Competence over time (retention and decay)
  - Integration
  - Transfer and generalization
- Performance classifications

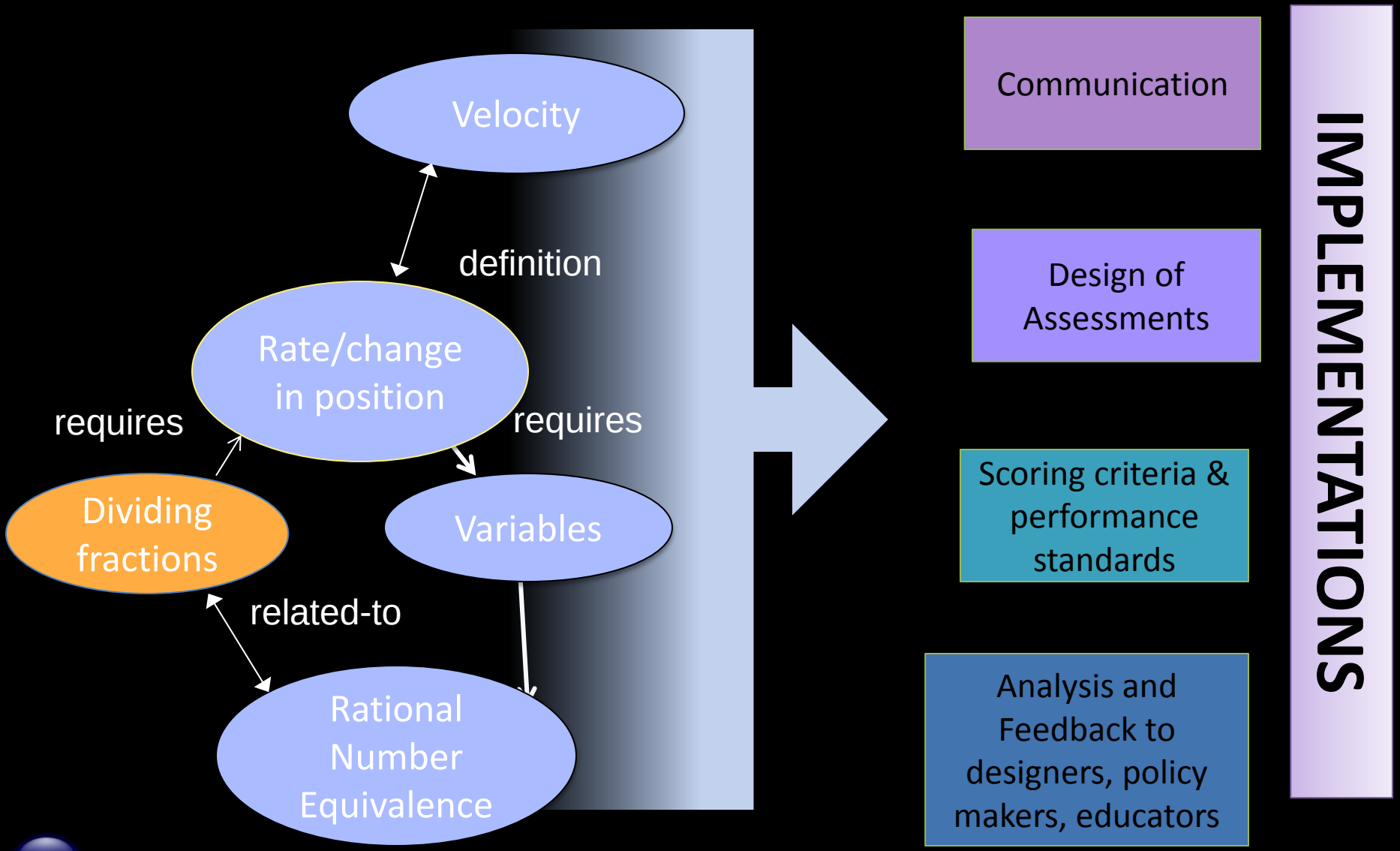
Tools to Help Design and Improvement of Assessments

- Content ontologies
- Cognitive (and other) ontologies
- Situations and fairness
- Combinatoric models for
  - Domain sampling
  - Qualitative attributes of complexity/difficult
  - Rapid generation of tasks
- Computational models for internal and external verification (validity)

Ontology Design



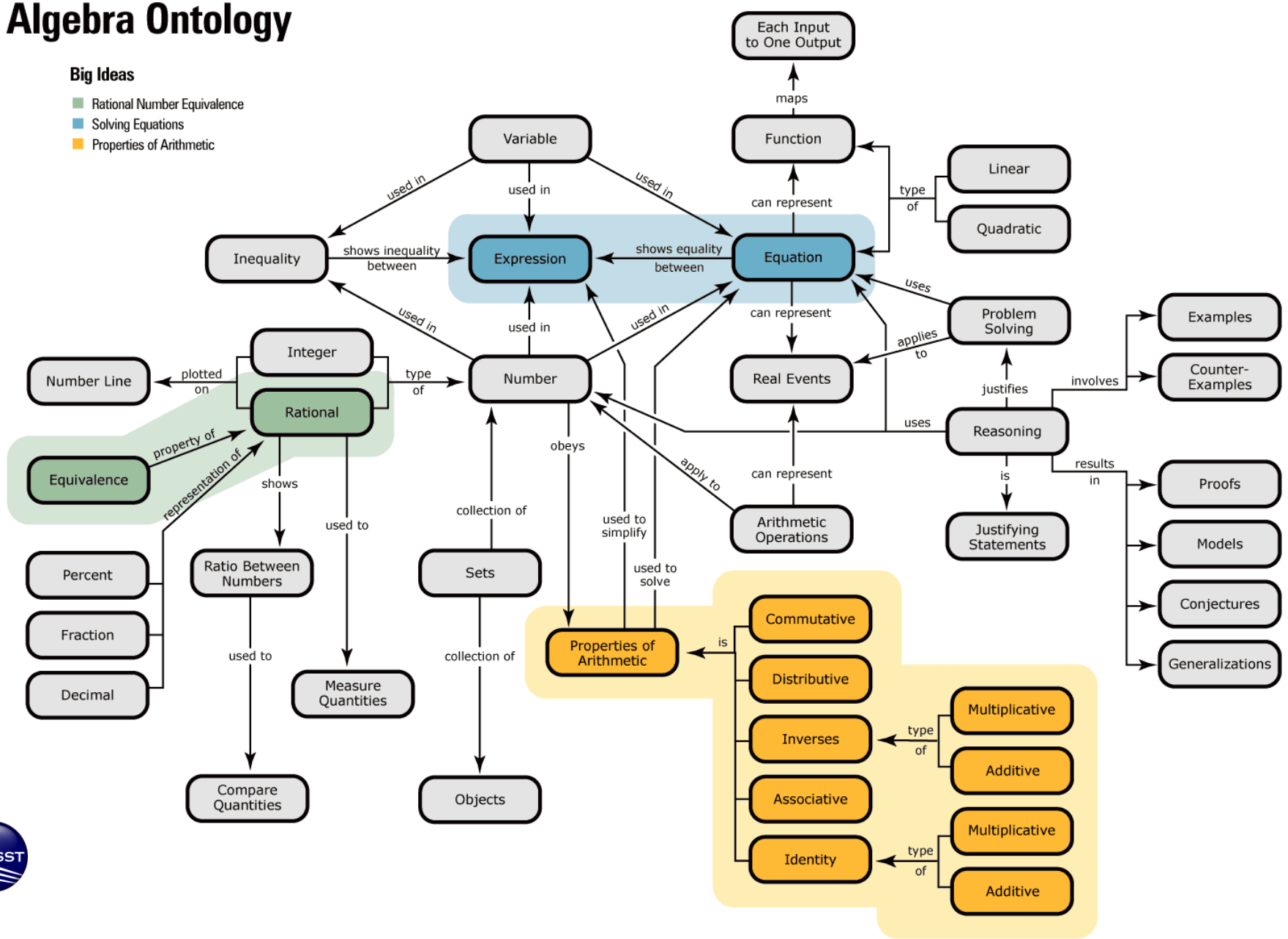
Ontology Use



Content Ontology

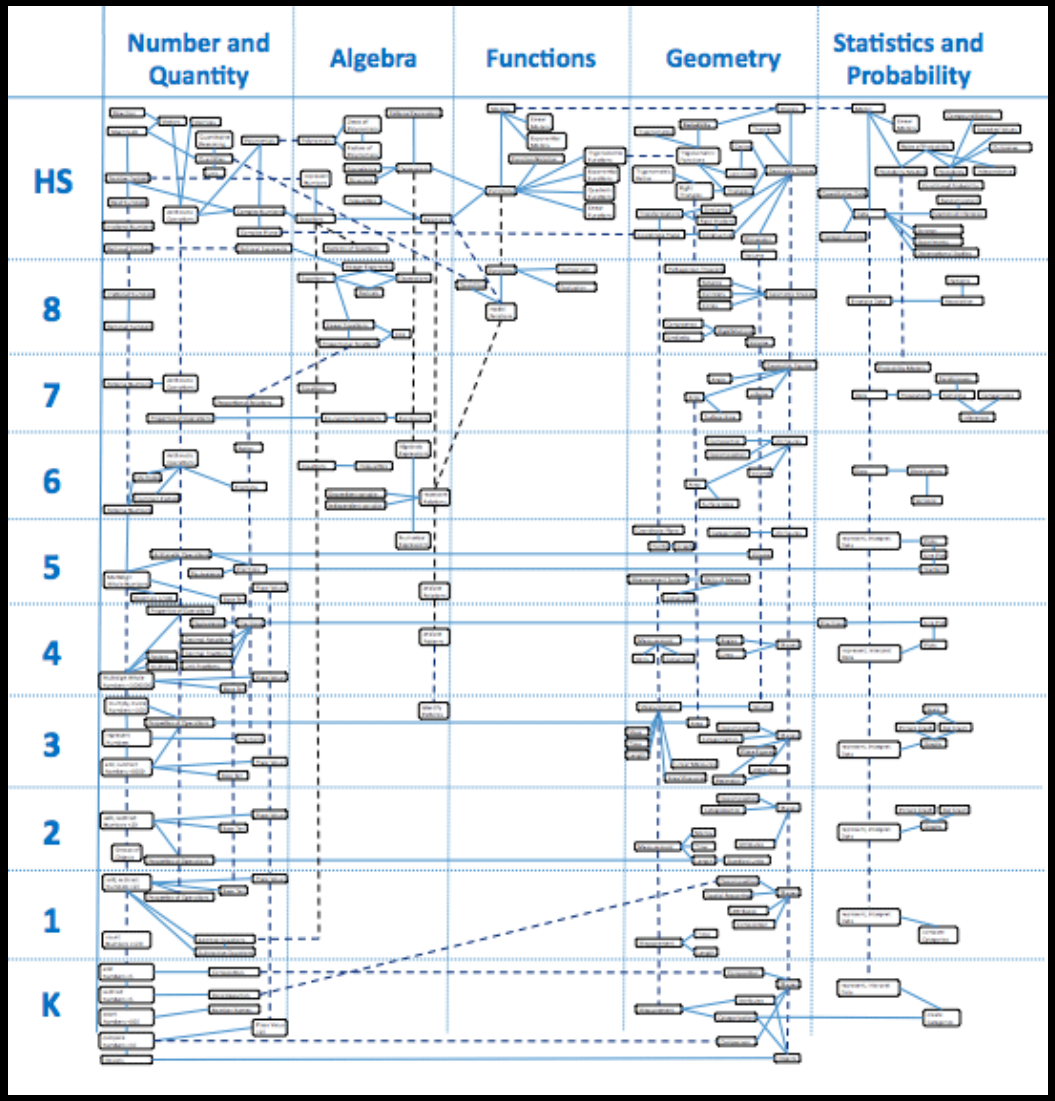
- Expert and Document Sources
  - Natural Language Processing
  - Experts
  - Computational form
- Used for sampling, comparison of interpretation of standards, instruction, tasks
- Gives structure to task & performance database
  - Cross sectional, longitudinal
- Crowdsourcing and data mining
- Instructional uses—sequencing, bird's eye view

Algebra Ontology

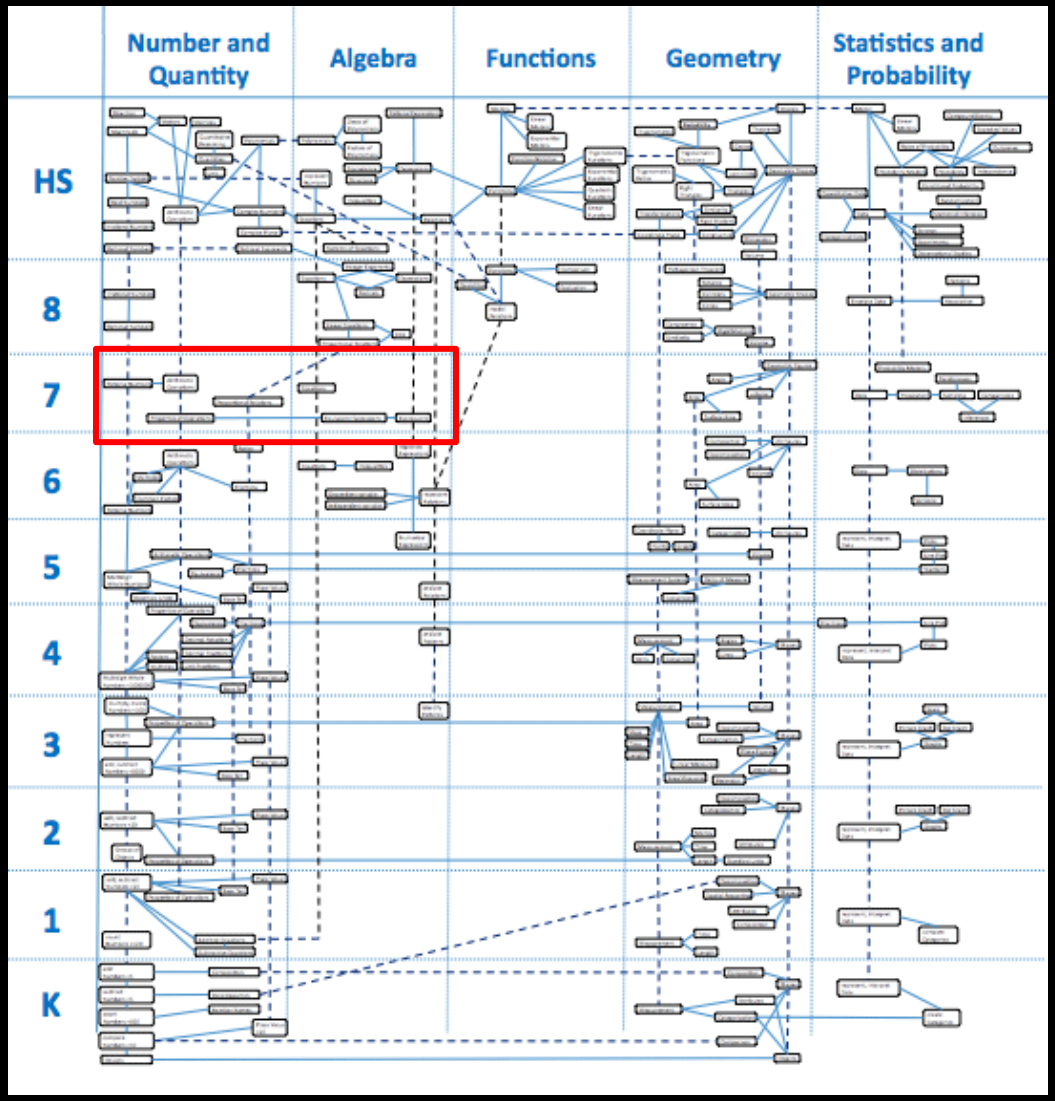


Niemi, D. © 2006 Regents of the University of California
Larry Casey - design

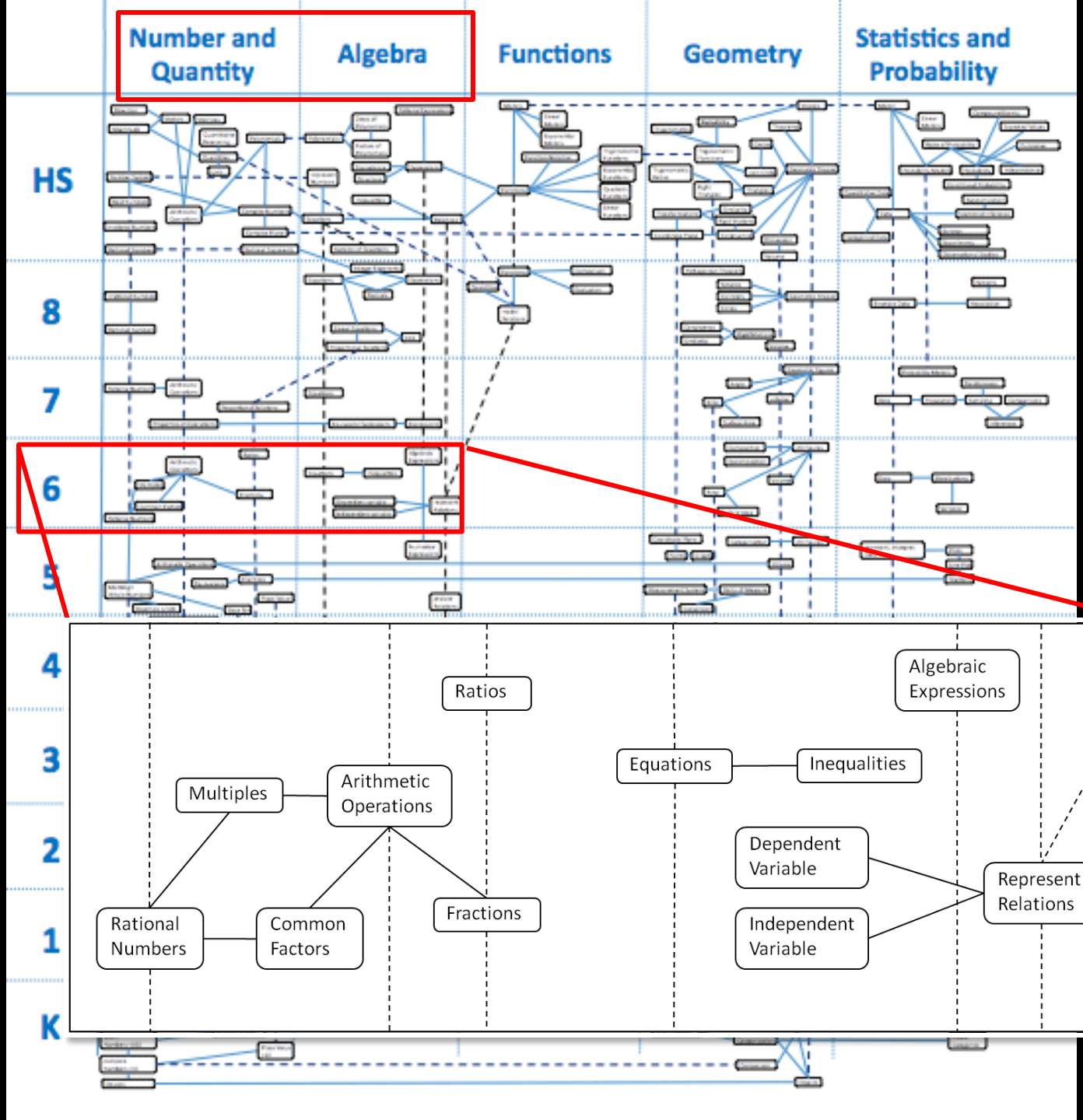
Common Core Ontology



Common Core Ontology



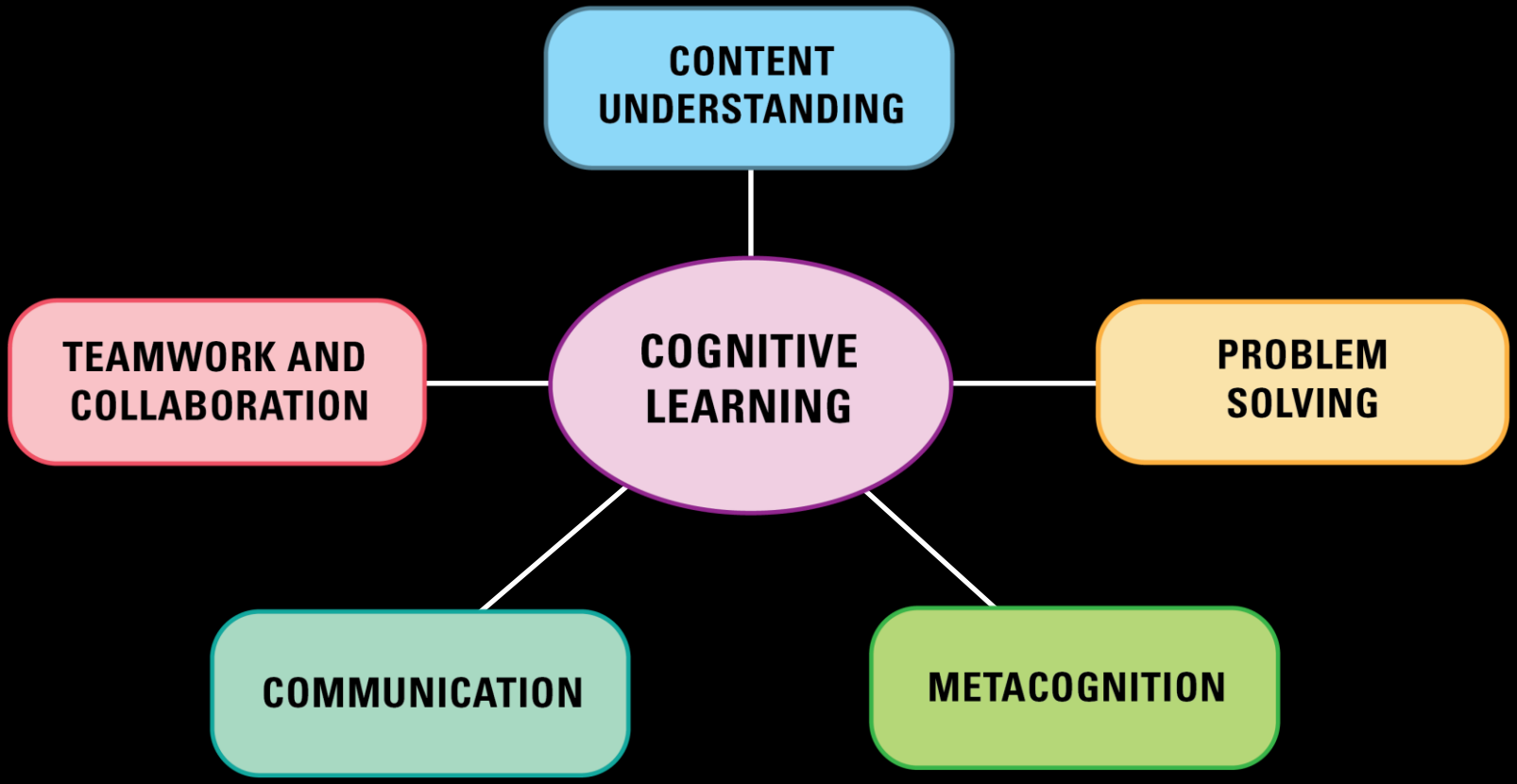
Common Core Ontology



Cognitive Demands, 21st Century Skills, Cognitive Readiness, Deeper Learning, Practices

- Purpose—sources of complexity
- Bound to content
- More than one may be combined
- Together with situational variables, prepare for unpredictability
- Individual differences vary, but focus is on learning and change, not talent

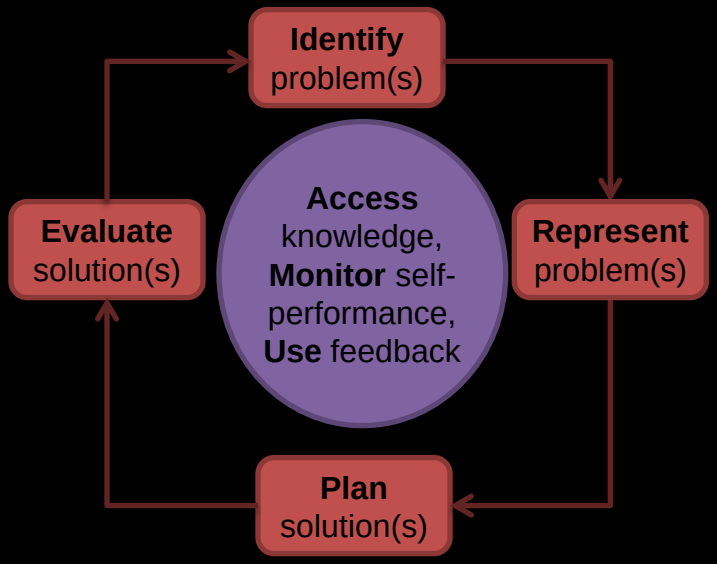
Cognitive Models for Assessment and Interventions



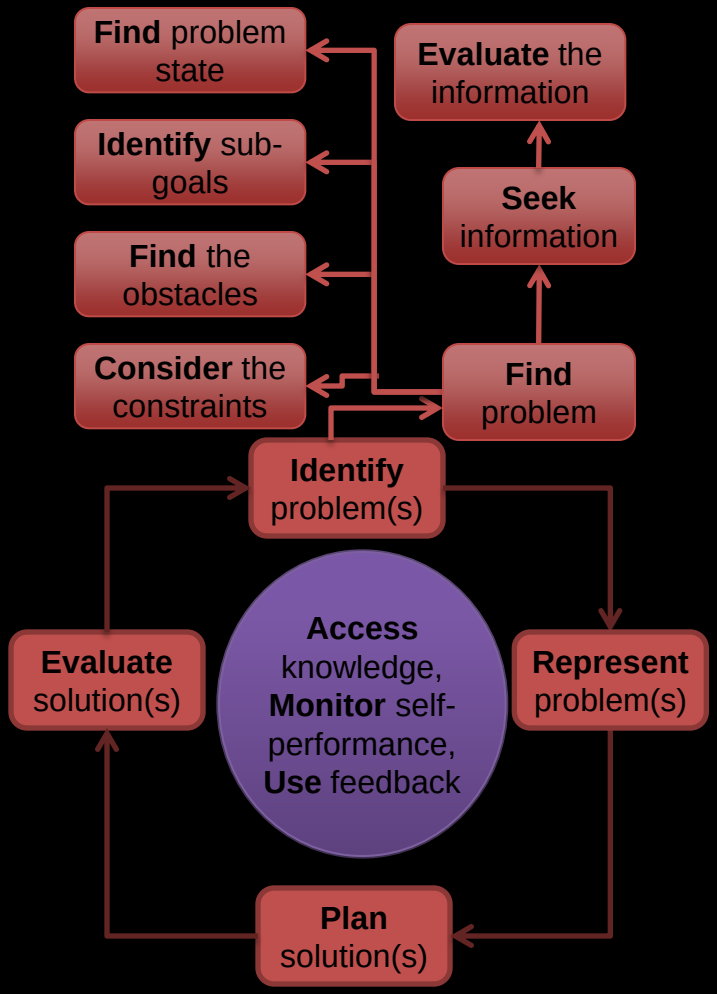
Expanded 21st Century Skills: Cognitive, Social, Intrapersonal

- Adaptive, complex problem solving
- Situation awareness and risk assessment
- Decision making
- Self-regulation – SEL
- Teamwork
- Learning to learn
- Communication
- Conceptual, procedural, and systemic learning of content

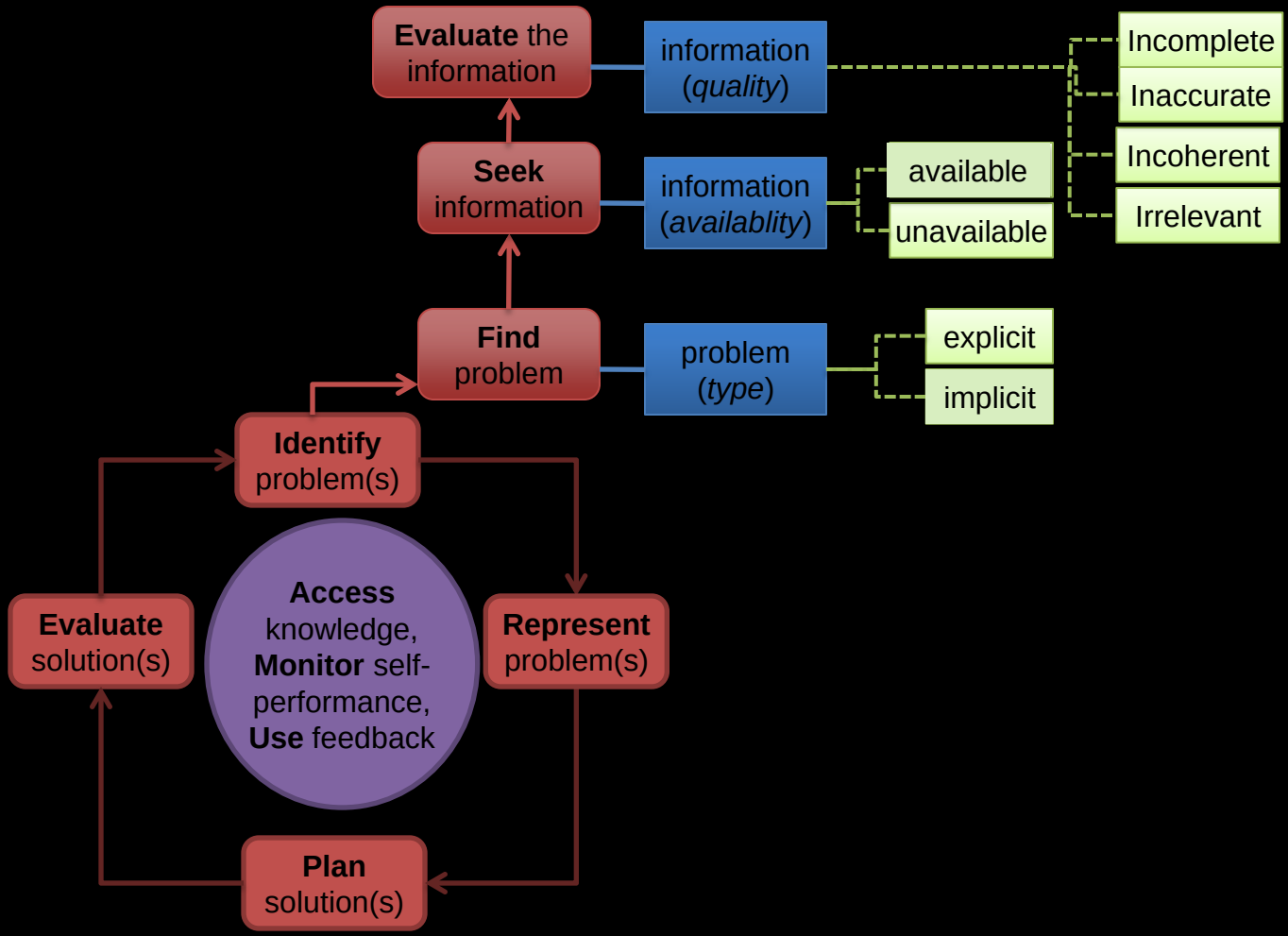
Problem Solving Ontology: *Student Processes*



Problem Solving Ontology: *Student Processes*

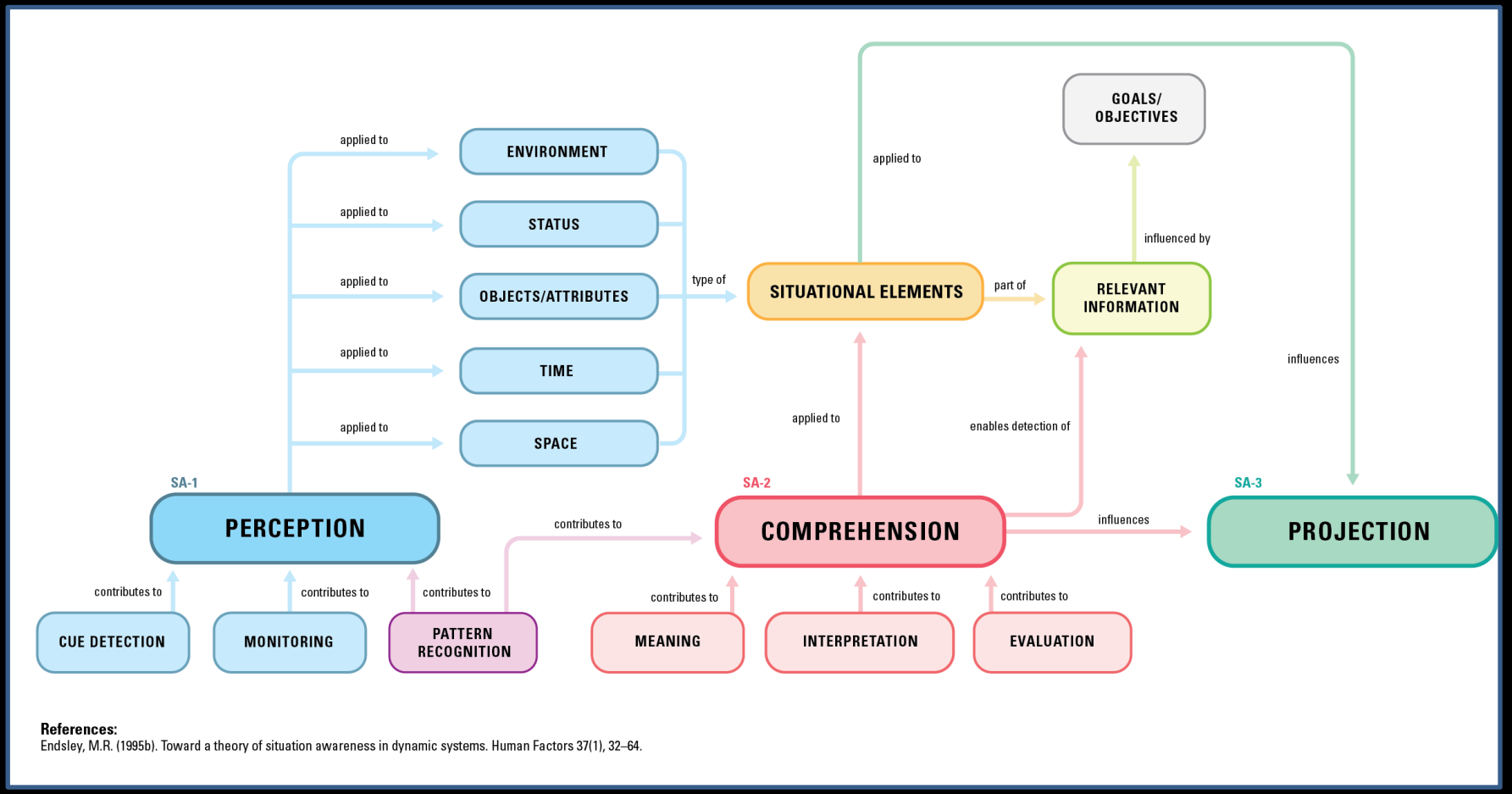


Problem Solving Ontology: *Problem Characteristics*



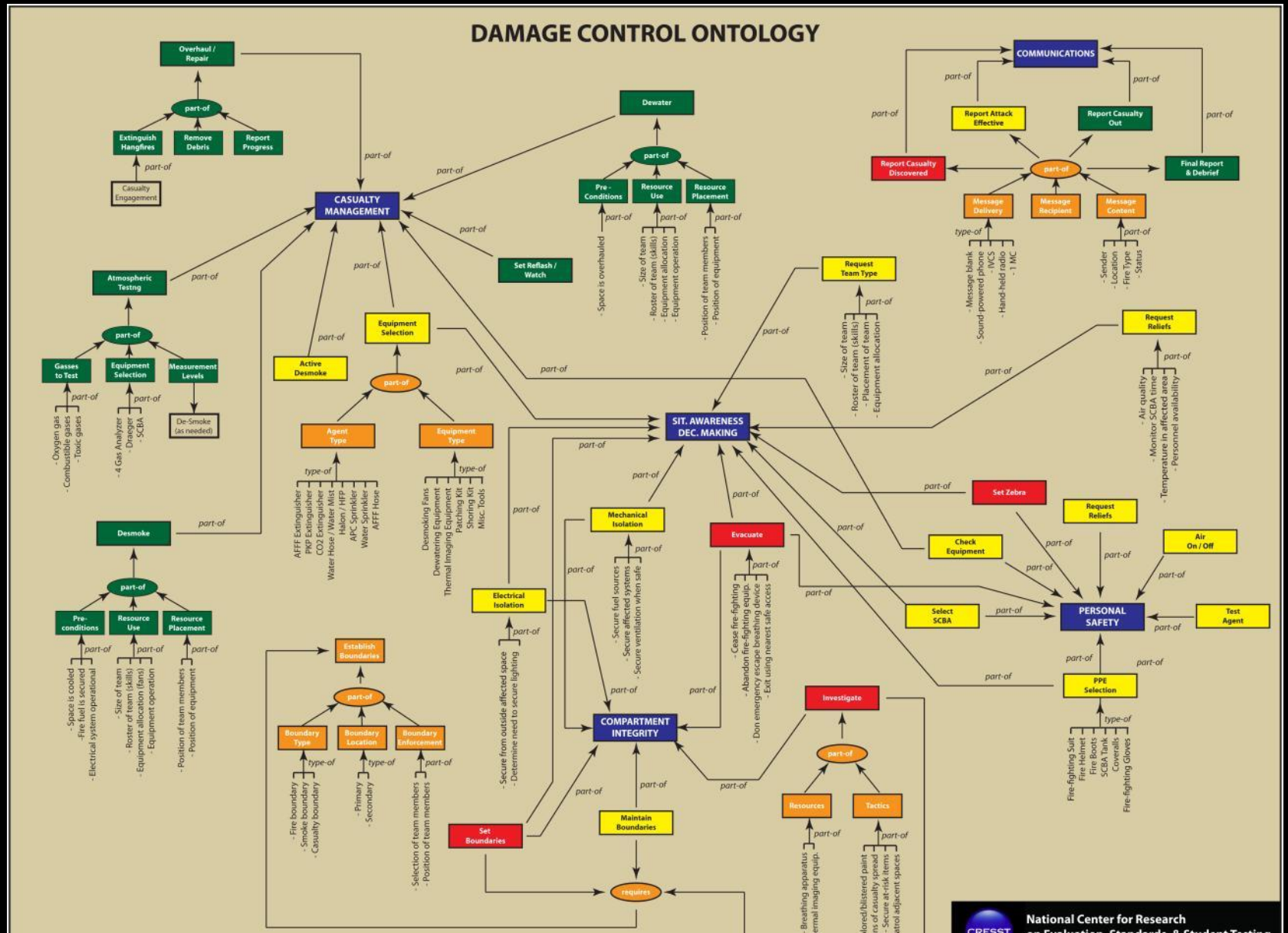
Problem-Solving Ontology–Designers

Situation Awareness

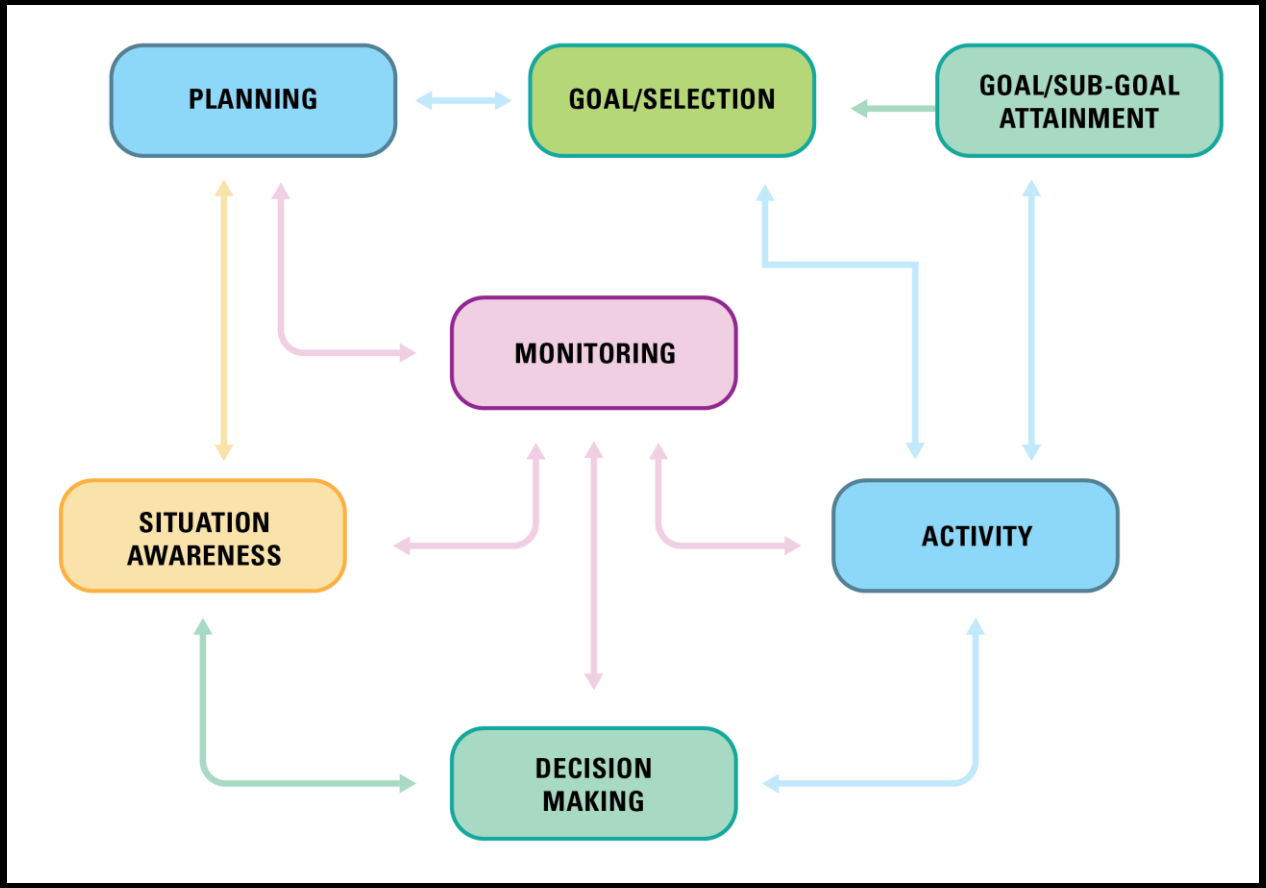


References:
Endsley, M.R. (1995b). Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), 32–64.

ONR VESSEL Situation Awareness & Communication



Metacognition

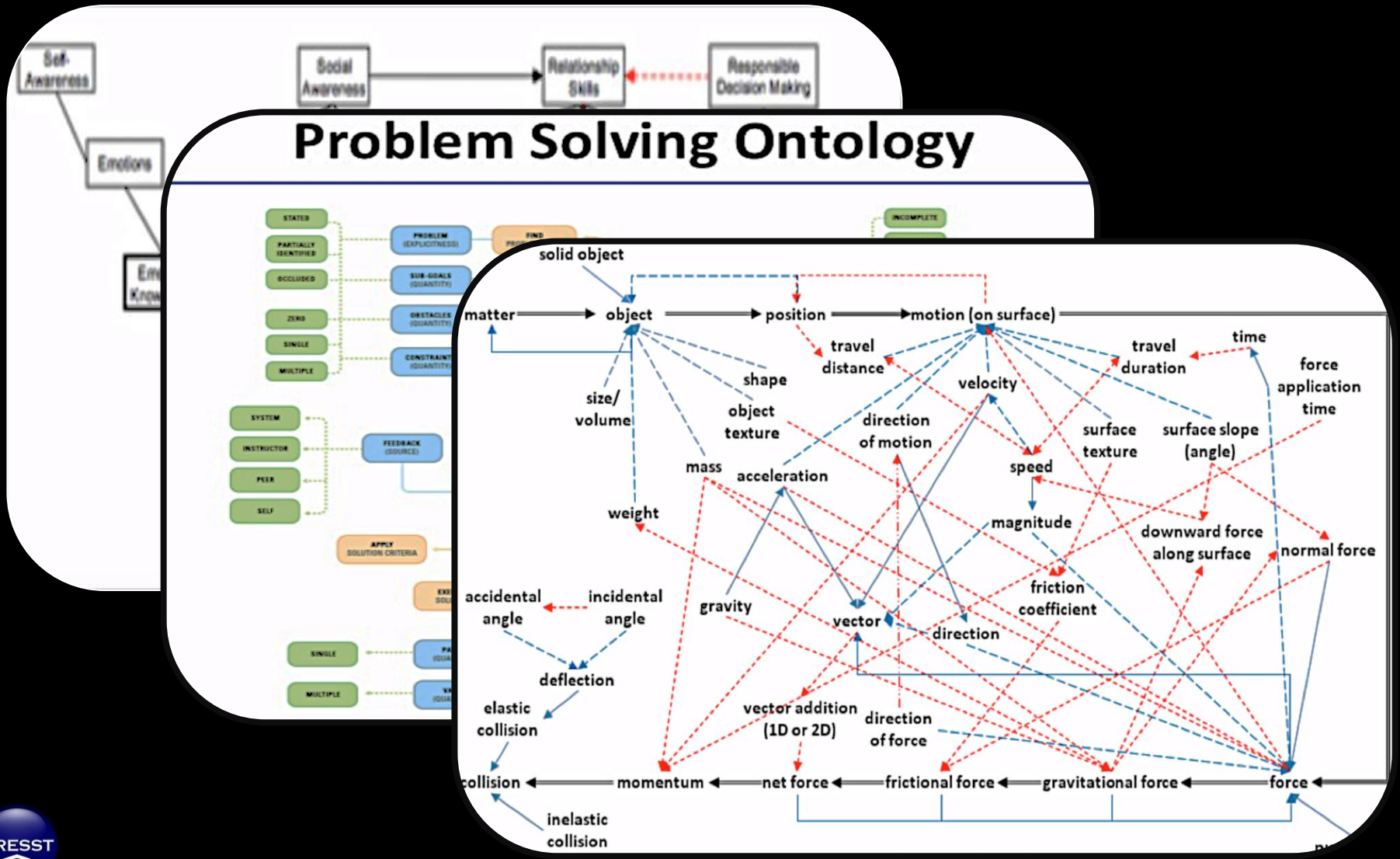


Combinatorics in Assessment Design

- Rules
- Features
  - Tasks situations, domain elements, prompts, help, constraints
  - Response requirements, e.g., process, solution, predictions, explanations, scoring
- For editing and curating
- Used with ontologies in games allows us to predict state and reason on series of formative assessments
- Provides feature based qualitative elements to use with computational models to explore validity under various conditions, e.g., students, instructional elements, sequence
- May solve the reliability issue involving performance tasks, that is comparablity

Tools to Use

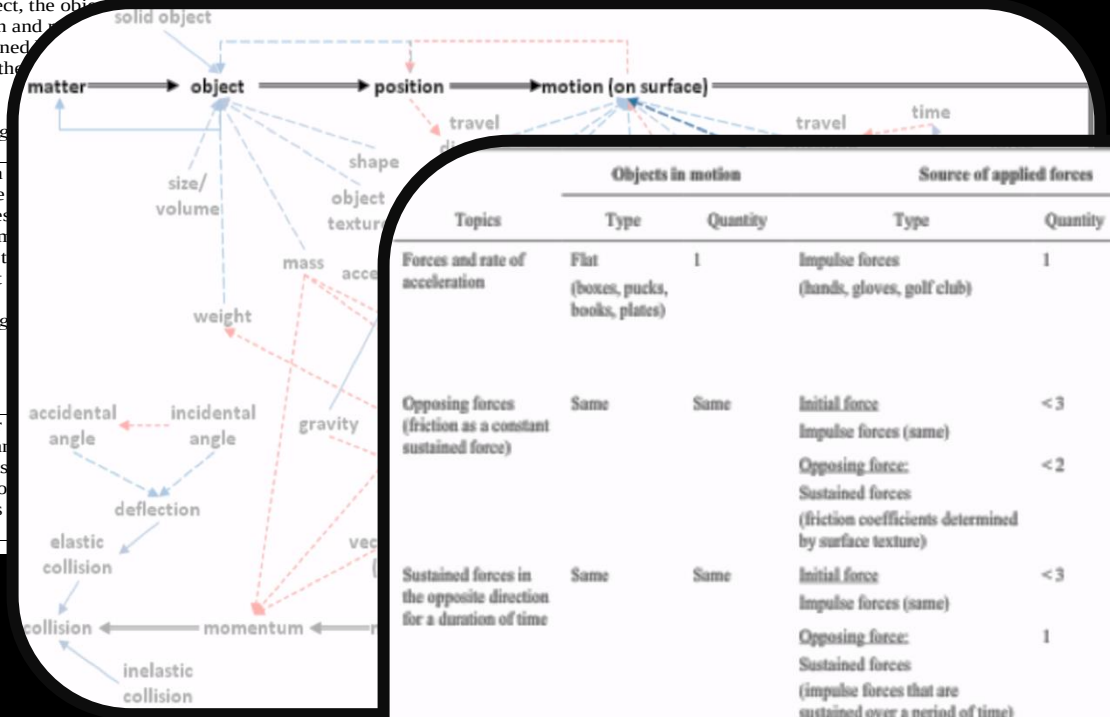
Ontologies (physics, socio-emotional learning, problem-solving)



Transformed into Formative/ Instructional Components

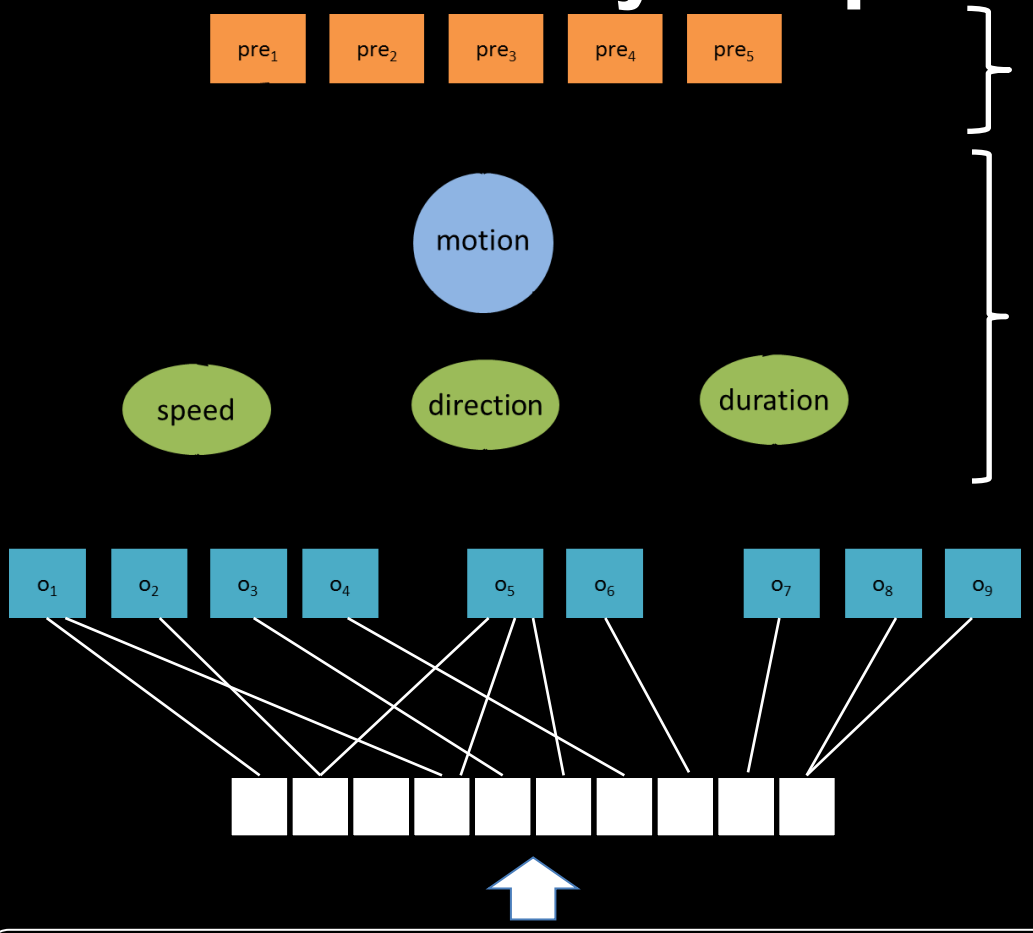
Instructional sequence and task specifications

Instructional Goals / ***Background Knowledge***	**Example Objectives**	**Physics Concepts Pertaining to Goals /** ***Misconceptions***
Students are able to explain that resultant forces arise when two or more forces in different directions are added together, and that when applied to an object, the obj… resulting direction of motion and … acceleration will be determined … magnitude and direction of the … (net) force applied. *Forces are composed of mag… direction.*	Students will apply multiple forces from different directions to the object to get the object to a specific	net force force as a vector vector addition (2D)
Students are able to model a … by applying a force opposite … of an object's motion, the res… slowing and/or stopping of m… depending on magnitude of t… force and duration of time it … *Forces are composed of mag… direction.*		
Students understand that for … motion (all moving at the sa… the more massive the objects … force in the opposite directio… required to bring the objects … given amount of time.		



Topics	Objects in motion: Type	Objects in motion: Quantity	Source of applied forces: Type	Source of applied forces: Quantity	Source of applied forces: Direction	Force magnitude: Categories	Force magnitude: Value
Forces and rate of acceleration	Flat (boxes, pucks, books, plates)	1	Impulse forces (hands, gloves, golf club)	1	Same	Easy (< 3) Hard (< 5)	Numerical (single digit integers) Small, medium, large Big, bigger, biggest Small, smaller, smallest
Opposing forces (friction as a constant sustained force)	Same	Same	Initial force Impulse forces (same)	< 3	Same	Same	Same
			Opposing force: Sustained forces (friction coefficients determined by surface texture)	< 2	Opposite		
Sustained forces in the opposite direction for a duration of time	Same	Same	Initial force Impulse forces (same)	< 3	Opposite	Same	Same
			Opposing force: Sustained forces (impulse forces that are sustained over a period of time)	1			
Velocity (vector addition)	Flat (boxes, pucks, books, plates) Round (balls, wheels)	Single	Impulse forces (same)	2	X-axis & Y-axis (Resultant force = 45° or 90°)	Same	Same

Validity–Computational Model



STUDENT BACKGROUND LAYER

- Prior knowledge, game experience
- Age, sex, language proficiency

CONSTRUCT LAYER

Construct, subordinate constructs, and inter-dependencies

INDICATOR LAYER

Behavioral evidence of construct, $o_n = f_n(q_1, q_2, q_3, \ldots q_n)$

Independent

FEATURE LAYER

Represents important features of the phenomena

$i_n = f_n(e_1, e_2, e_3, \ldots; s_1, s_2, s_3, \ldots)$: Computes a feature value given raw events and states

Transformation Functions

Extracts feature

events and states (e_1, e_2, e_3, ...; sB_1, s_2, s_3, ...) Be

EVENT LAYER

Raw behavior and states

Dependent

Validity Minimums

- Validation of measures
  - Expert review of the alignment to the target ontologies
  - Think-aloud protocols to determine whether expected learner concepts, procedures, practices, processes are being applied
  - Identification and evaluation of critical paths of performance
  - Evidence of instructional sensitivity
  - Psychometric studies of reliability, dimensionality and fairness (total scores and diagnostic subscales)
  - Evidence obtained before wide application

Summary- Our Collaboration

- Your job is to translate this into your own rich knowledge of mathematics
- Our (technical community) job is to develop quality standards to support the credibility, validity, and usability of assessments

http://www.cse.ucla.edu

Eva L. Baker
voice 310.206.1530
email eva@ucla.edu

Citations:

p. 3: http://c2.ac-images.myspacecdn.com/images02/129/l_0157a4bec6d648a482fafb71d28befa9.jpg
p. 6:
http://www.google.com/search?hl=en&site=imghp&tbm=isch&source=hp&biw=1003&bih=565&q=boy+with+pencil+in+hand&oq=boy+with+pencil&gs_l=img.1.1.0l2.2374.6386.0.9673.17.11.1.5.5.0.105.912.10j1.11.0...0.0...1ac.1.8.img.i1WEwepTyC4#hl=en&site=imghp&tbm=isch&sa=1&q=black+boy+writing&oq=black+boy+writing&gs_l=img.12..0j0i24.94947.101120.0.103828.21.15.1.4.4.1.300.1311.13j3-1.14.0...0.0...1c.1.8.img.PSTWKNOryGA&bav=on.2,or.r_qf.&bvm=bv.44697112,d.cGE&fp=83bce1ba5cf84439&biw=1003&bih=565&imgrc=ICcf_0GWMthNlM%3A%3Bu-UW7eWIxeyMbM%3Bhttp%253A%252F%252Fwww.thenorthstarnews.com%252FContent%252F101%252Fblack%252520boy%252520studying.jpg%3Bhttp%253A%252F%252Fwww.thenorthstarnews.com%252FIssue.aspx%253FID%253D2011-05-19%3B183%3B276;http://library.thinkquest.org/04oct/00492/Quiz_pencil.jpg;
http://casualgamerchick.com/2007/06/30/achievements-coming-to-casual-games/