

Assessment and Evaluation of Outcomes in Engineering Education

James W. Pellegrino

UIC LEARNING SCIENCES
RESEARCH INSTITUTE

Foci of Today's Presentation

- 21st Century Skills – What are they? Why so much interest? Relevance to STEM education?
 - Relevant background ideas about the domain and the implications for instruction and assessment
- Examples drawn from Engineering Education
 - Assessment of conceptual understanding
- Assessment in the Context of Program Evaluation
- Some Implications for Assessment and Evaluation in Engineering Education

EDUCATION

AT THE NATIONAL RESEARCH COUNCIL
www.nationalacademis.org/dbasse

**EDUCATION
FOR LIFE
AND WORK**

Developing Transferable
Knowledge and Skills in
the 21st Century

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

Committee on Defining Deeper Learning and 21st Century Skills

Division of Behavioral and Social
Sciences and Education
National Research Council

Study & Report Context

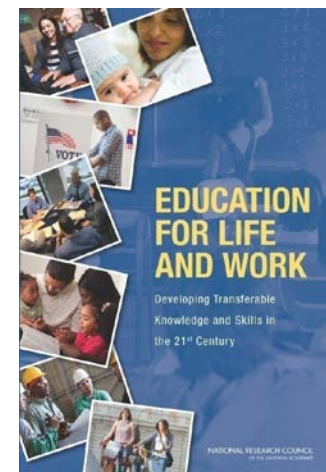
- Education is a shared endeavor, including schools, teachers, nonprofit groups, informal learning institutions, taxpayers, parents, and the students themselves.
- Business and educational leaders are asking this shared endeavor to infuse development of “21st century skills” such as problem solving, critical thinking, and collaboration into teaching and learning.
- A variety of names are used to refer to these skills.
- To help the public understand the research related to these skills, several foundations charged the U.S. NRC....

Committee Charge

- Define the set of key skills referred to as “deeper learning,” “21st century skills,” and by other labels
- Review research on their importance for positive adult outcomes
- Describe how the skills relate to each other and to the learning of reading, mathematics, and science and engineering
- Discuss how to teach and assess them and the implications for teacher education and professional development

Clarifying Terms & Identifying Domains of Competence

UIC LEARNING SCIENCES
RESEARCH INSTITUTE



Skills Identified in an Influential OECD Survey

1. Creativity/innovation
2. Critical thinking
3. Problem solving
4. Decision making
5. Communication
6. Collaboration
7. Information literacy
8. Research and inquiry
9. Media literacy
10. Digital citizenship
11. Information and communications technology operations and concepts
12. Flexibility and adaptability
13. Initiative and self-direction
14. Productivity
15. Leadership and responsibility
16. Other (please specify)

Many, if not all, of these
are what we expect of
the graduates from our
engineering programs

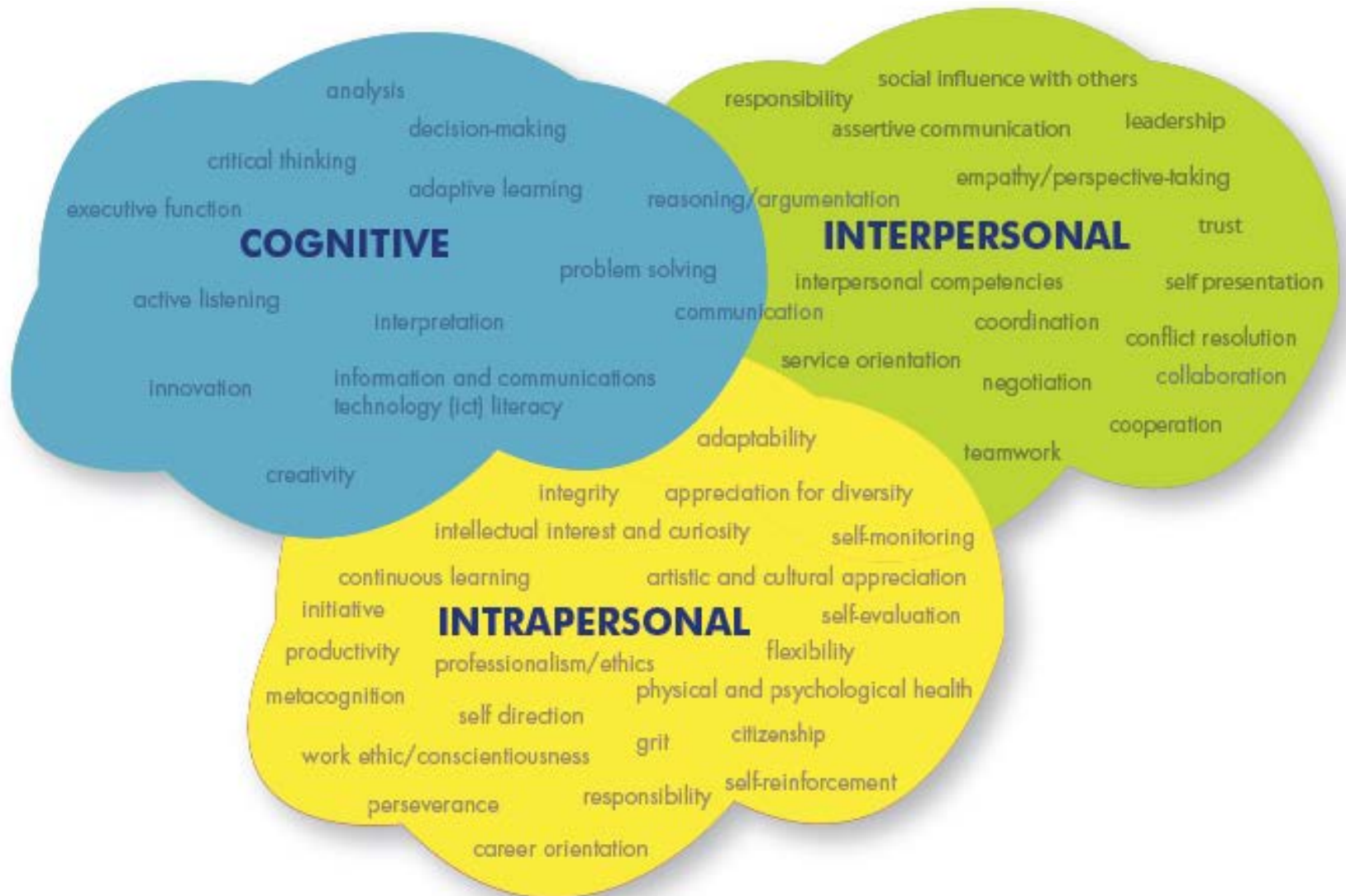
Source: Adapted from Ananiadou, and Claro (2009).

Clarifying Terms

- Deeper learning is the process of learning for transfer. It enables an individual to take what was learned in one situation and apply it to new situations.
- The product of deeper learning is transferable knowledge, including content knowledge in a subject area and procedural and strategic knowledge of how, why, and when to apply this knowledge to answer questions and solve problems in the subject area.
- We refer to this transferable knowledge as “21st century competencies” to reflect that both skills and knowledge are included.



Three Domains of Competence

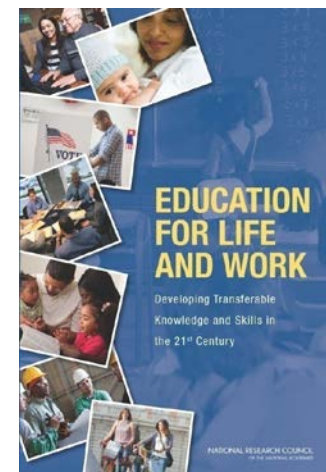


Three Domains of Competence: Competency Clusters

- The **Cognitive Domain** includes three clusters of competencies:
 - cognitive processes and strategies
 - knowledge
 - creativity
 - These clusters include competencies such as critical thinking, **problem solving**, information literacy, reasoning and argumentation, and innovation.
- The **Intrapersonal Domain** includes three clusters of competencies:
 - intellectual openness
 - work ethic and conscientiousness
 - positive core self-evaluation
 - These clusters include competencies such as flexibility, initiative, appreciation for diversity, and metacognition (the ability to reflect on one's own learning and make adjustments accordingly).
- The **Interpersonal Domain** includes two clusters of competencies:
 - teamwork and collaboration
 - leadership
 - These clusters include competencies such as communication, **collaboration**, responsibility, and conflict resolution.

Evidence of Importance for Each Competency Domain

UIC LEARNING SCIENCES
RESEARCH INSTITUTE

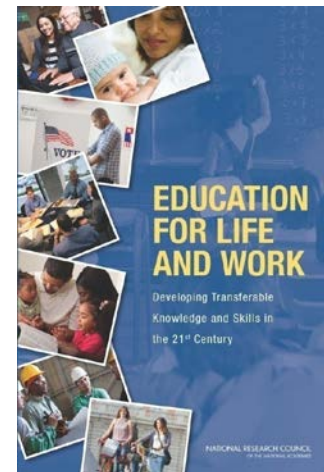


Evidence of Importance for Adult Outcomes

- Cognitive competencies show modest positive correlations with desirable educational, career, and health outcomes.
- Achievement, especially math achievement, predicts most strongly to labor market outcomes
- Conscientiousness correlates most strongly with job performance
- Anti-social behavior is often predictive as well

Teaching for Transfer & Deeper Learning in the Disciplines

UIC LEARNING SCIENCES
RESEARCH INSTITUTE



Teaching for Transfer

- Emerging evidence indicates that cognitive, intrapersonal and interpersonal competencies can be taught and learned in ways that promote effective transfer.
- Transfer is not domain general – it is domain specific.



Transfer is Supported When:

- Learners understand general principles, as emphasized in the recent U.S. standards in mathematics, science and English language arts.
- Learners understand factual and conceptual knowledge in a subject area and also applicable problem-solving strategies.
- Learners recognize how, when, and why to apply their factual, conceptual, and procedural knowledge and skills.



To Design Instruction for Transfer:

- Begin with clearly-defined learning goals and a model of how learning is expected to develop.
- Use assessments to measure and support progress toward goals.
- Provide multiple, varied representations of concepts and tasks.
- Encourage questioning and discussion.
- Engage learners in challenging tasks, with support and guidance.
- Teach with carefully selected sets of examples and cases.
- Prime student motivation.
- Use formative assessment to provide feedback.



Foci of Today's Presentation

- 21st Century Skills – What are they? Why so much interest? Relevance to STEM education?
 - Relevant background ideas about the domain and the implications for instruction and assessment
- **Examples drawn from Engineering Education**
 - **Assessment of conceptual understanding**

A Recurring Engineering Education Scenario

- Established sets of topics are commonly taught in STEM courses with a focus on developing content knowledge and practical problem solving skills
- This often results in students able to pass course exams without achieving deep conceptual understanding within the discipline

A Recurring Engineering Education Dilemma

- STEM instructors regularly lament their students' difficulties in developing deep conceptual understanding within their particular STEM disciplines
- They simultaneously lament the challenges they face in defining and effectively assessing student understanding.
- They want tools, methods & data

What we ostensibly seek: “Informative” Assessments

- Assessments designed to measure and report on student knowledge and understanding in a manner and format that can directly support teaching and learning processes in the classroom.
 - Useful and “useable” information
- These can include assessments embedded in curricula as well as more distal assessments.
- The information derived from the assessments can be used to support a formative assessment process.
 - Cycle of feedback and deployment can vary



NATIONAL RESEARCH COUNCIL

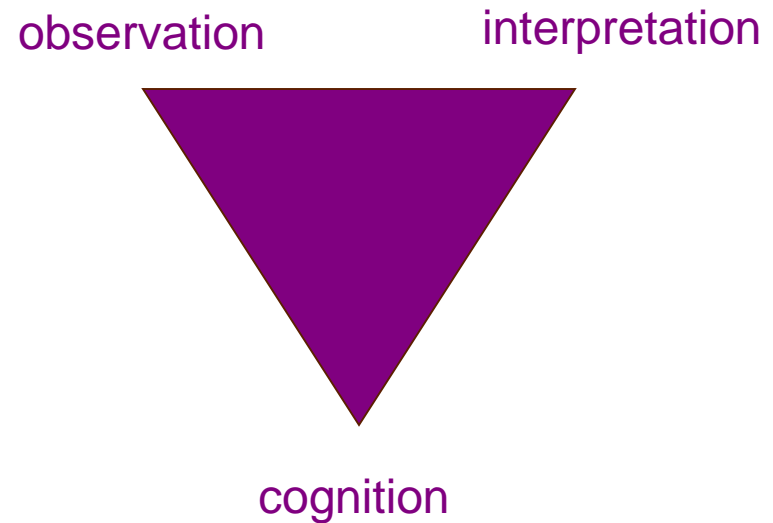


Knowing what Students Know

The Science
and Design
of Educational
Assessment

Assessment as a Process of Reasoning from Evidence

- cognition
 - theory or model of how students represent knowledge & develop competence in the domain
- observations
 - tasks or situations that allow one to observe students' performance
- interpretation
 - method for making sense of the data



*Must be
coordinated!*

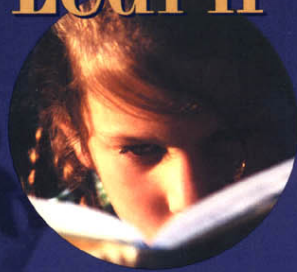
Scientific Foundations of Educational Assessment

- Advances in the Sciences of Thinking and Learning -- **the cognition vertex**
 - informs us about what observations are important and sensible to make
- Contributions of Measurement and Statistical Modeling -- **the interpretation vertex**
 - Informs us about how to make sense of the observations we have made

Why Models of Development of Domain Knowledge are Critical

- Tell us what are the important aspects of knowledge that we should be assessing.
- Give us strong clues as to how such knowledge can and should be assessed
- Can lead to assessments that yield more instructionally useful information
 - diagnostic & prescriptive
- Can guide the development of systems of assessments
 - work across contexts & time

How People Learn



Brain,



Mind,



Experience,

and

School

NATIONAL RESEARCH COUNCIL

How Students Learn



**HISTORY,
MATHEMATICS,
AND SCIENCE
IN THE
CLASSROOM**

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

Journal of Engineering Education

THE RESEARCH JOURNAL FOR ENGINEERING EDUCATION

SPECIAL ISSUE
EDUCATING FUTURE ENGINEERS:
WHO, WHAT, AND HOW

- 227 **A Rising Global Discipline**
Jack R. Lohmann
- 231 **On Becoming a 21st Century Engineer**
Sheri D. Sheppard, James W. Pellegrino,
and Barbara M. Olds
- 235 **Special Guest Editorials**
Charles M. Vest, Shirley M. Malcom,
Robin S. Adams, and Richard M. Felder
- 241 **Sponsor Commentaries**
Xavier Fouger, Ray Almgren, Kris Gopalakrishnan,
and Paul Mailhot
- 245 **Educating Generation Net—Can U.S.
Engineering Woo and Win the
Competition for Talent?**
Daryl Chubin, Krista Donaldson, Barbara Olds,
and Lorraine Fleming
- 259 **Persistence, Engagement, and Migration
in Engineering Programs**
Matthew W. Ohland, Sheri D. Sheppard,
Gary Lichtenstein, Ozgur Eris, Debbie Chachra,
and Richard A. Layton
- 279 **Learning Conceptual Knowledge in the
Engineering Sciences: Overview and
Future Research Directions**
Ruth A. Streveler, Thomas A. Litzinger,
Ronald L. Miller, and Paul S. Steif
- 295 **Looking Beyond Content: Skill
Development for Engineers**
Edward F. Redish and Karl A. Smith
- 309 **Characterizing Design Learning:
A Mixed-Methods Study of Engineering
Designers' Use of Language**
Cynthia J. Atman, Deborah Kilgore,
and Ann McKenna
- 327 **Ethics Teaching in Undergraduate
Engineering Education**
Anne Colby and William M. Sullivan
- 339 **Conceptualizing Engagement:
Contributions of Faculty to Student
Engagement in Engineering**
Helen L. Chen, Lisa R. Lattuca,
and Eric R. Hamilton
- 355 **Becoming an Engineer: Toward a Three
Dimensional View of Engineering Learning**
Reed Stevens, Kevin O'Connor, Lari Garrison,
Andrew Jaccus, and Daniel M. Amos
- 369 **Advancing Engineering Education
in P-12 Classrooms**
Sean Brophy, Stacy Klein, Merredith Portsmore,
and Chris Rogers
- 389 **The Academic Bookshelf**
Diane T. Rover



July 2008 | Vol. 97 No. 3



Aditya Johri
Barbara M. Olds

CAMBRIDGE HANDBOOK OF

Engineering Education Research

CAMBRIDGE

Engineering Education & Conceptual Understanding

- Developing deep conceptual knowledge is difficult
- How can engineering educators assess whether students:
 - understand key concepts?
 - have propensities for particular misconceptions?
- Concept Inventories (CIs)
 - Potential to assess conceptual understanding
 - But... to what extent are they *valid* and *reliable* indicators of student knowledge?

General Features of CIs

- CIs typically assess a relatively narrow domain: “the concept of force” in physics (FCI, Hestenes); the area of “statics” (CATS, Steif & Dantzler); or “heat transfer, thermodynamics, and fluid mechanics” (TTCI, Streveler, Olds, Miller, Nelson)
- CIs attempt to measure deeper conceptual understanding, not just rote facts or procedures
- CIs typically are used in courses in high school, college, community college, & technical schools

Claims often made about CI's

- They measure deep conceptual understanding
- They deliberately avoid questions that require computation or use of formulas
- Their multiple choice format provides for scoring efficiency & objectivity
- Multiple choice distractors are explicitly linked to common student errors or misconceptions
- CI scores can be used to evaluate the impact of alternative instructional approaches

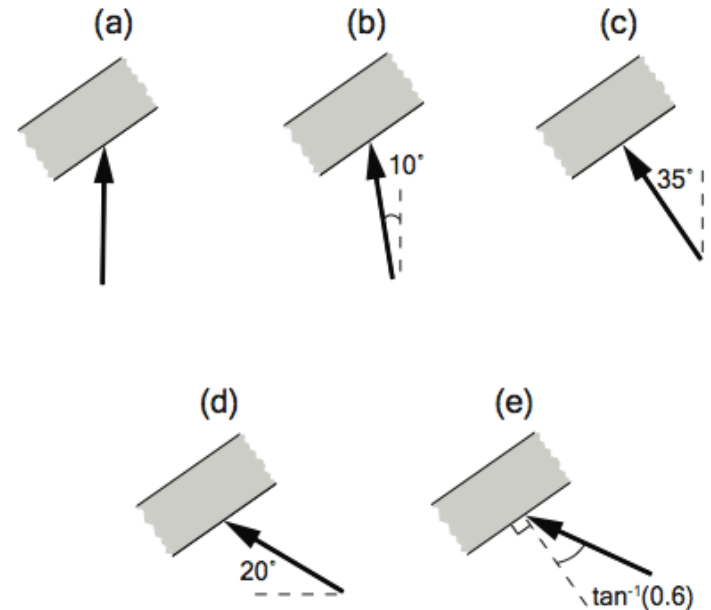
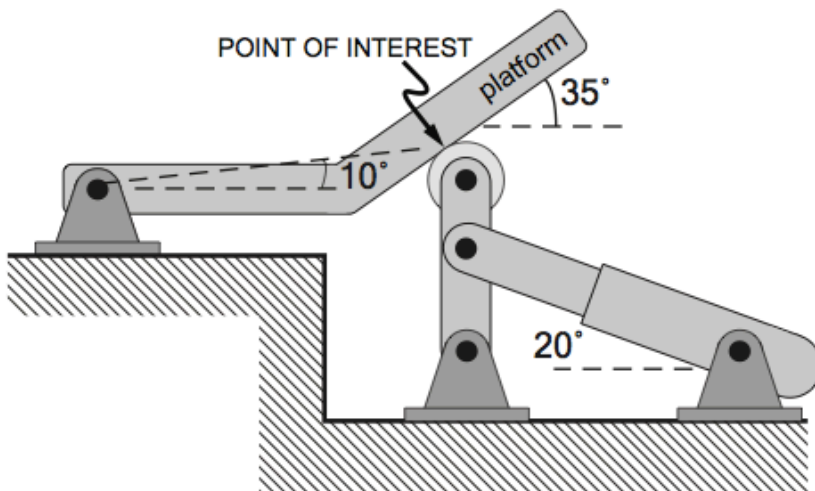
Examples of Specific CIs We Have Studied

- CATS -- Statics, Paul Stief
- TTCl -- Heat Section, Ron Miller and Ruth Streveler
- DCI – Dynamics, Paul Gray

Concept Assessment Tool for Statics (CATS) Example

11. The platform is kept in the position shown by a roller, link and hydraulic cylinder. The pin that the roller rotates on is frictionless. The coefficient of friction between the roller and the platform is 0.6.

What is the direction of the force exerted by the roller on the platform at the point of interest?



Thermal & Transport Inventory (TTCI) Example

Two identical closed beakers contain equal masses of liquid at a temperature of $20\text{ }^{\circ}\text{C}$ as shown below. One beaker is filled with water and the other beaker is filled with ethanol (ethyl alcohol). The temperature of each liquid is increased from $20\text{ }^{\circ}\text{C}$ to $40\text{ }^{\circ}\text{C}$ using identical heaters immersed in the liquids. Each heater is set to the same power setting.

It takes 2 minutes for the ethanol temperature to reach $40\text{ }^{\circ}\text{C}$ and 3 minutes for the water temperature to reach $40\text{ }^{\circ}\text{C}$.



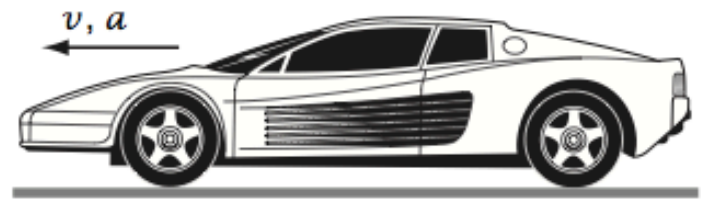
Ignoring evaporation losses, to which liquid was more energy transferred during the heating process?

- Water because it is heated longer
- Alcohol because it heats up faster (temperature rises faster)
- Both liquids received the same amount of energy because they started at the same initial temperature and ended at the same final temperature
- Can't determine from the information given because heat transfer coefficients from the water and alcohol beaker surfaces are needed
- Can't determine from the information given because heat capacities of water and ethanol are needed
- Water because it has a higher boiling point than ethanol.

Dynamics Concept Inventory (DCI) Example

Questions 21 & 22

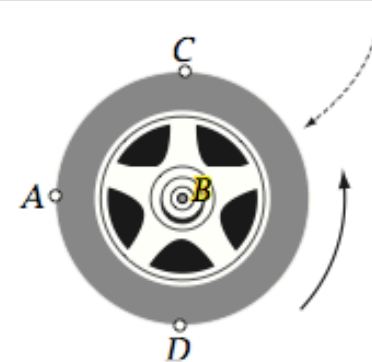
A rear-wheel drive car moves with speed v and acceleration a to the left as shown in the figure. The tires do not slip on the road.



Question 21

What is the direction of the velocity of point A attached to the tire?

- (a) ← (b) ↙ (c) ↓ (d) ↘ (e) ↑



Question 22

Of the points A , B , C , and D shown on the tire, which has the smallest speed?

- (a) A (b) B (c) C (d) D (e) They all have the same speed.

Assessment Validity

- Validity: extent to which an assessment measures what it was intended to measure
 - Dependent on proposed interpretation and use of test scores
- Evidence related to validity is demonstrated both in analyses of the assessment's content and of examinee response patterns

Assessment Validation as an Evidence-Based Argument

- Establish the intended ***claims*** about student understanding based upon aspects of student performance
- Indicate what ***evidence*** would support these claims
- Multiple forms of evidence should be collected and interpreted to determine the extent to which they support the intended assessment claims

Evidentiary Validity Framework: Evaluating Claims for CIs

Typical claims about interpretive uses of a concept inventory.

CI scores can be used to indicate students'

1. Overall mastery of *all* concepts represented in the CI

2. Mastery of particular sets of concepts

3. Propensity for misconceptions or common errors

The Evidentiary Validity Framework

- An analytic framework that CI researchers and test developers can use to evaluate the validity of an existing CI relative to its specific interpretive claims
- Each claim has associated data analytic methods that provide evidence for evaluating a particular type of claim
 - Overall score – item functioning & test reliability
 - Sub-scores – scale reliability & internal coherence
 - Error patterns – interpretability & consistency

Applying the Framework to Two Concept Inventories

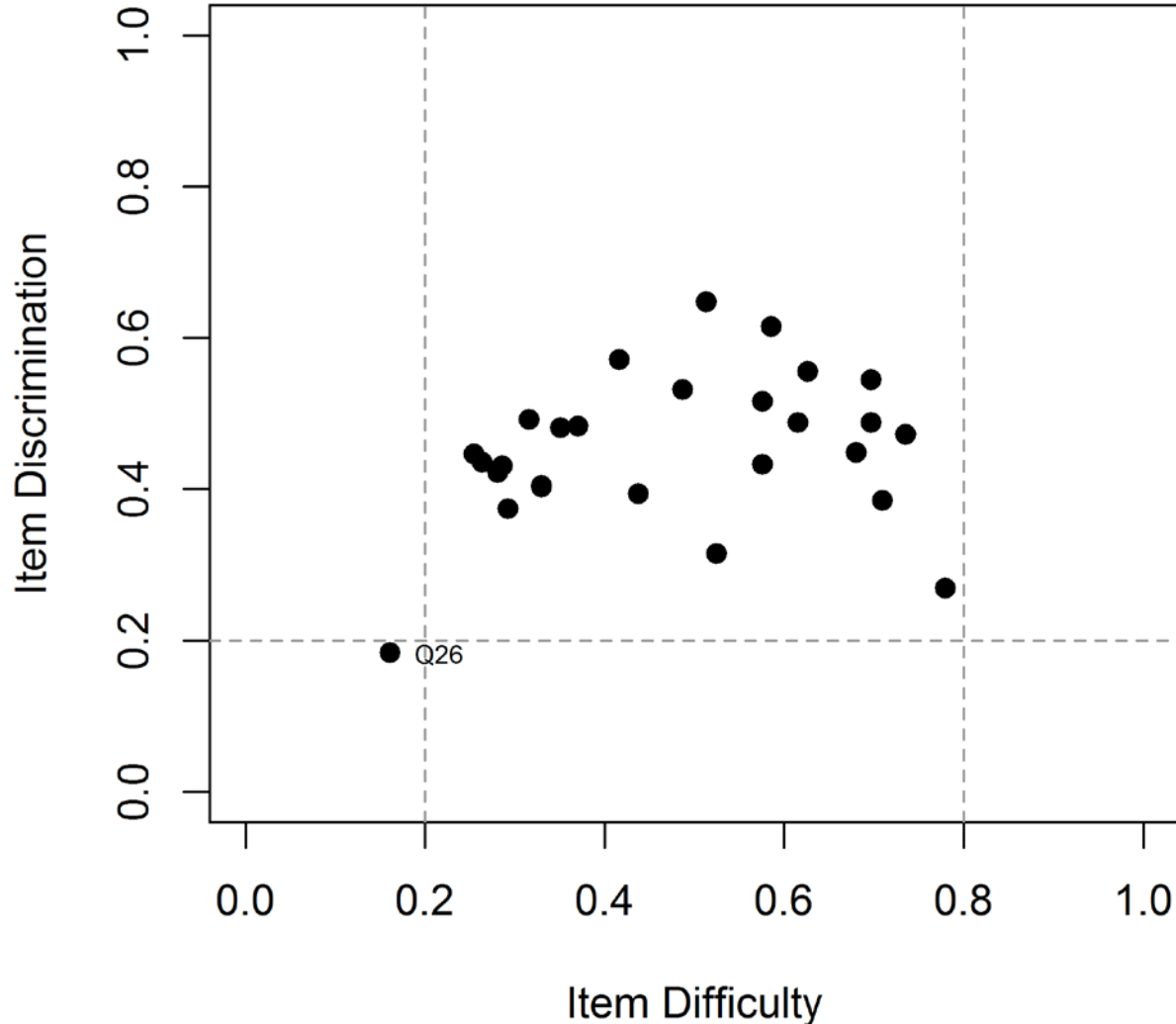
1. Dynamics Concept Inventory (DCI)

- Gray, Costanzo, Evans, Cornwell, Self, & Lane (2005)
- 29 items; 11 categories
- Dataset: $N = 996$ undergraduates

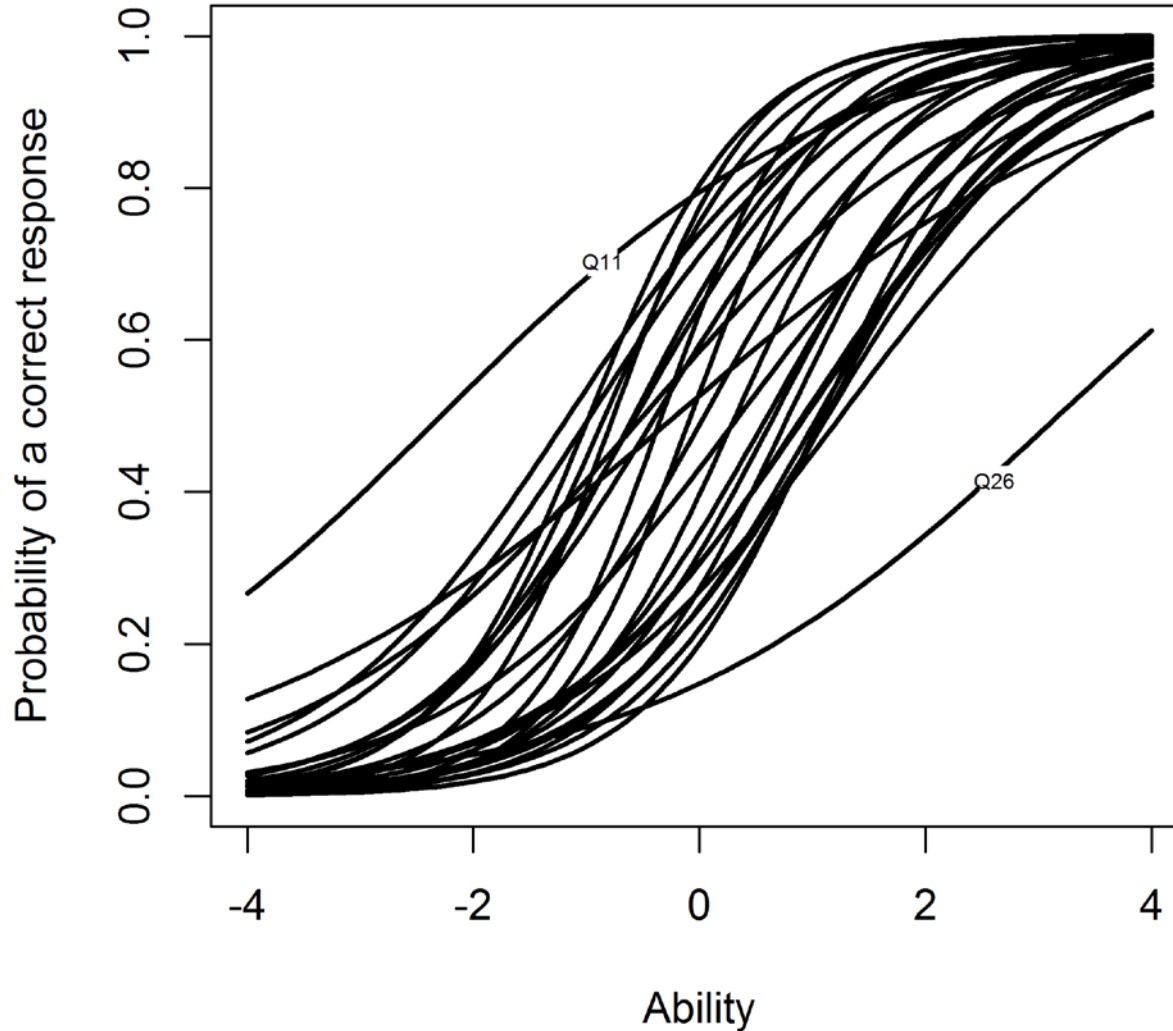
2. Concept Assessment Tool for Statics (CATS)

- Steif & Dantzler (2005)
- 27 items; 9 categories
- Dataset: $N = 1372$ undergraduates

CATS: Claim 1 – Overall Domain Mastery



CATS: Claim 1 – Overall Domain Mastery



Summary – Evaluating Evidence for CI Claims

- Claim 1 – Overall Domain Mastery
 - For both DCI & CATS, majority of items performed well. Consider dropping (replacing) problematic items before using total score. Measurement error quantified.
- Claim 2 – Mastery of Particular Concepts
 - DCI subscores are problematic – low internal consistency; weak match to hypothesized categories
 - CATS subscores are useful – internally consistent subscales; close match to hypothesized categories
- Claim 3 – Identify misconceptions/errors
 - Neither DCI nor CATS produced evidence to support claim about clear identification of misconceptions
- Learn more in a forthcoming issue of the *Journal of Engineering Education*

Conclusions & Applications

- For CI scores to be meaningfully interpreted, developers and users of these instruments need to consider issues of validity
- Our Evidentiary Validity Framework provides guidelines for evaluating validity arguments for CIs.
 - Useful for existing CIs (with student data). Can suggest areas to focus redesign efforts.
 - Can inform development of novel CIs. Explicated claims can guide item and instrument development.

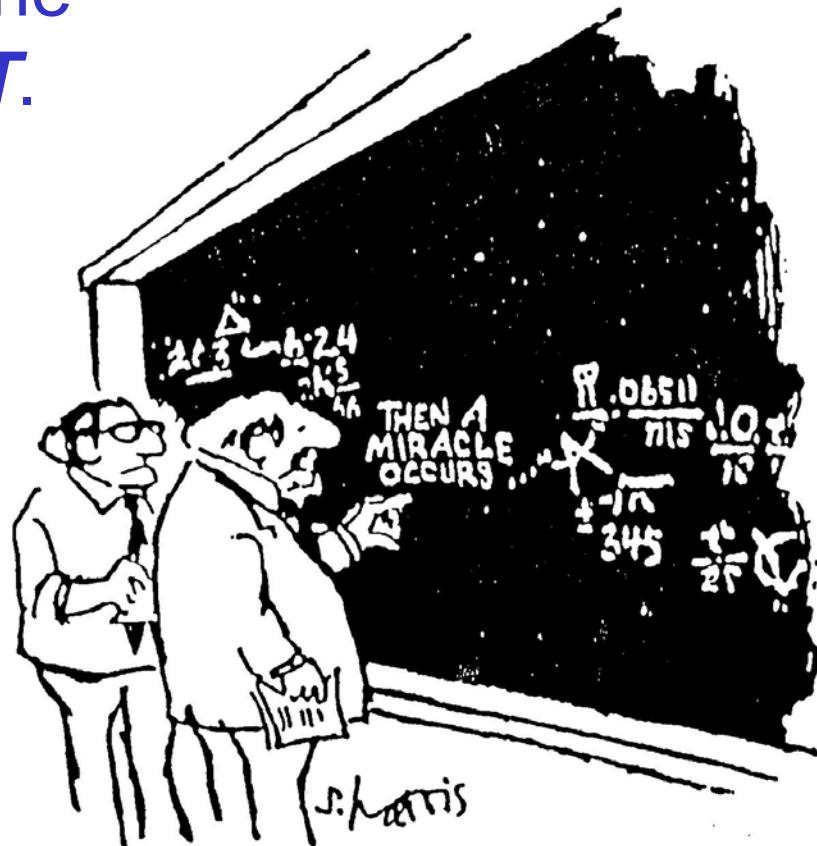
Foci of Today's Presentation

- 21st Century Skills – What are they? Why so much interest? Relevance to STEM education?
 - Relevant background ideas about the domain and the implications for instruction and assessment
- Examples drawn from Engineering Education
 - Assessment of conceptual understanding
- **Assessment in the Context of Program Evaluation**

A common problem is that a Program's activities and strategies may or may not lead to the desired outcomes.

A logic model makes the connections ***EXPLICIT***.

"I think you should be more explicit here in Step Two."



A logic model is...

- A depiction of a program showing what the program does and what it intends to accomplish.
- A series of “if-then” relationships that, if implemented as intended, ostensibly lead to the desired outcomes
- The core of program planning and evaluation

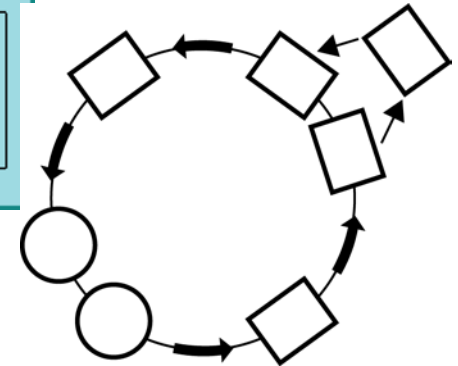
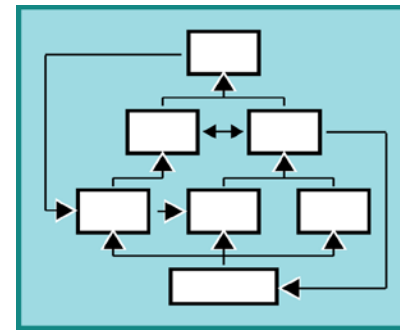
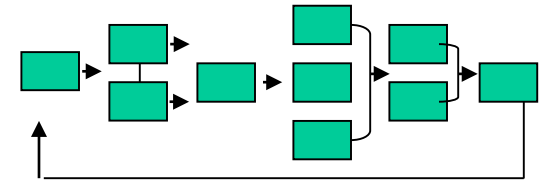
Logic modelling is a way of thinking...
not just a pretty graphic

"We build the road and the road builds us."
-Sri Lankan saying



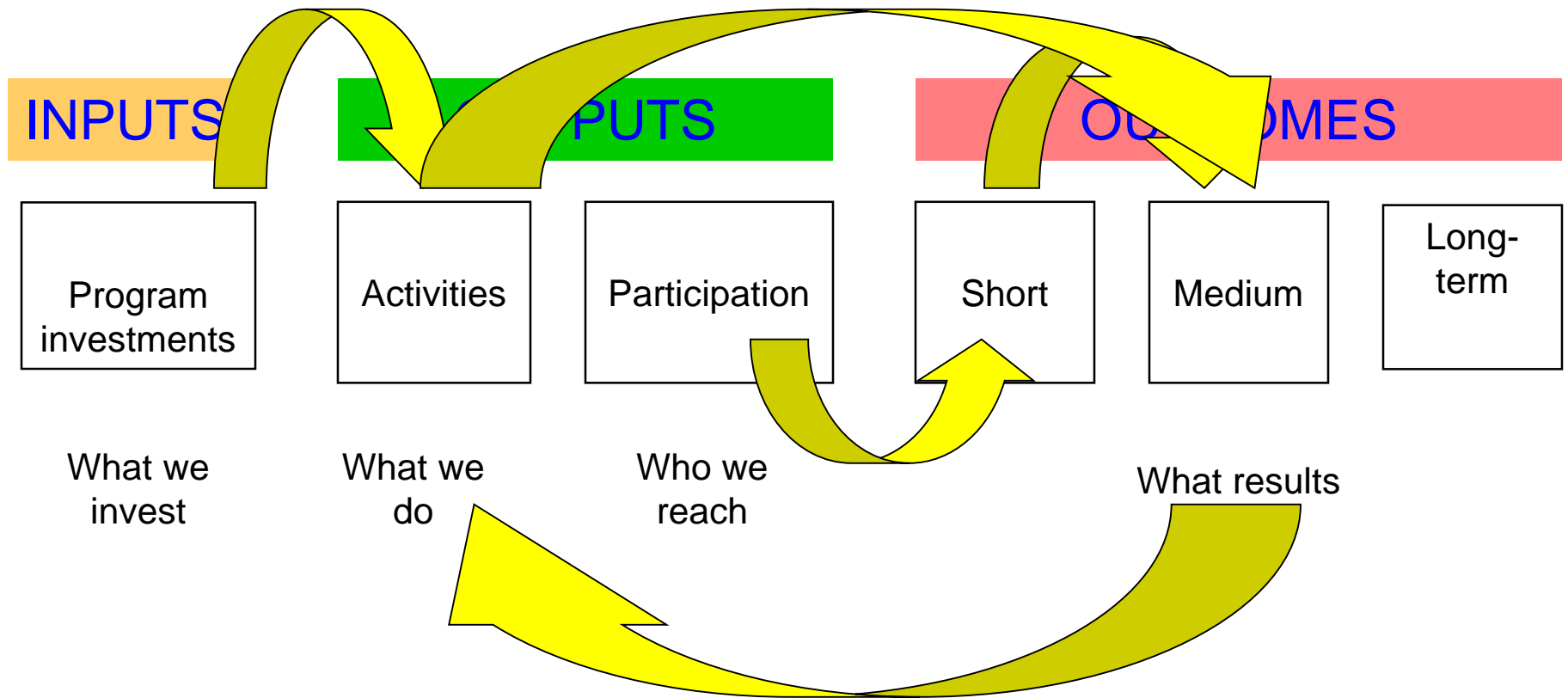
What does a logic model look like?

- Graphic display of boxes and arrows; vertical or horizontal
 - Relationships, linkages
- Any shape possible
 - Circular, dynamic
 - Cultural adaptations; storyboards
- Level of detail
 - Simple
 - Complex
- Multiple models
 - Multi-level programs
 - Multi-component programs



Inputs	Outputs	Outcomes
	1	1a b
	2	
	3	2a b c
	4	3a b

Feedback loops and multi-dimensions



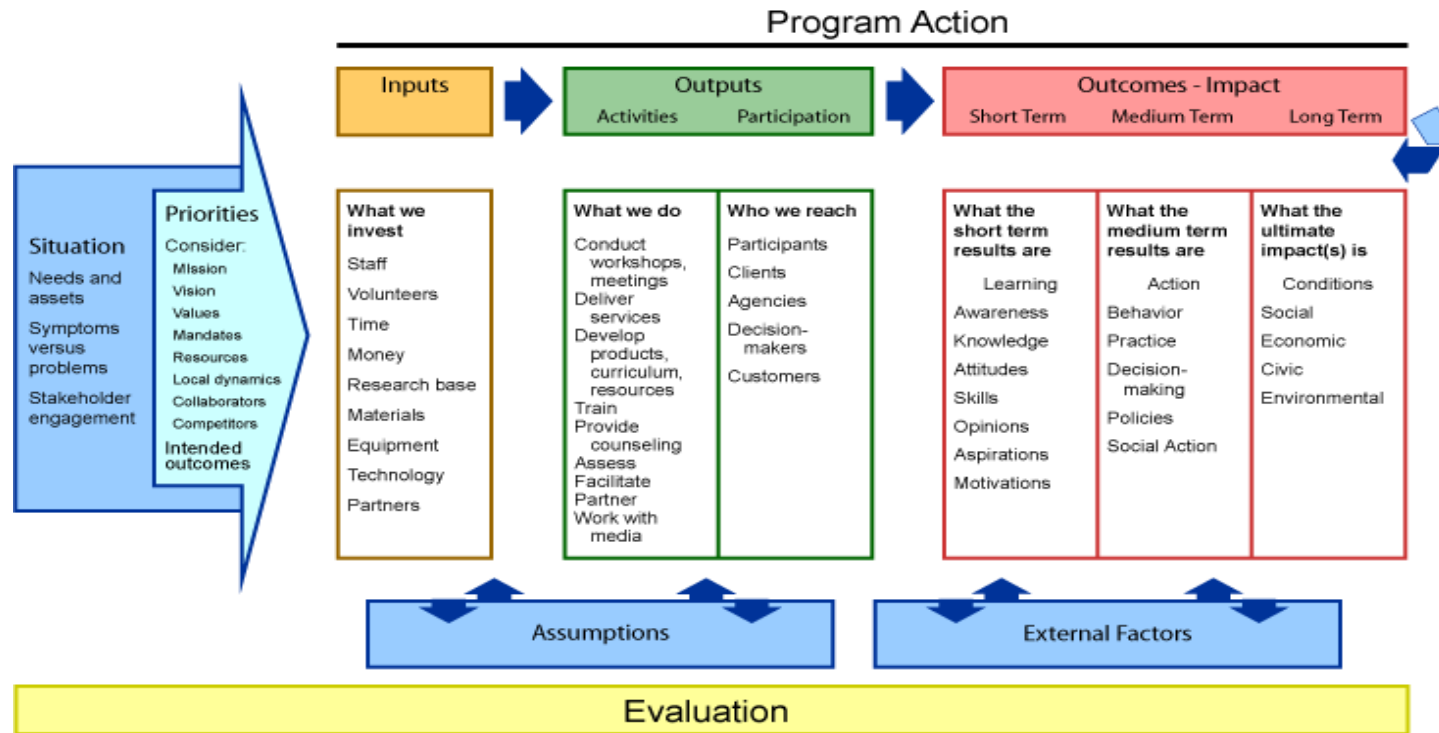
What a logic model is not...

- A theory
- Reality
- An evaluation model or method

It is a framework for describing the relationships between investments, activities, and results.

It provides a common approach for integrating planning, implementation, evaluation and reporting.

Logic model in evaluation



What do you want to know?

How will you know it?

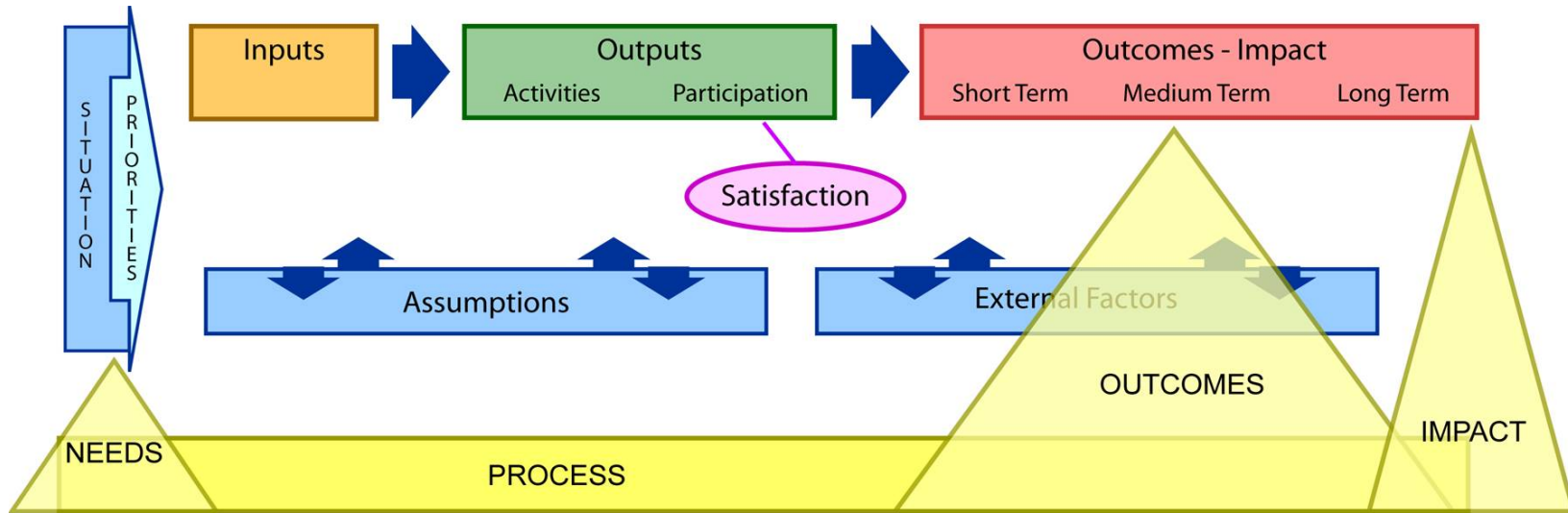
EVALUATION: check and verify

Logic Model helps with Evaluation

Provides the program description that guides an evaluation process

- Helps you match evaluation to the program
- Helps you know what and when to measure
 - Are you interested in process and/or outcomes?
- Helps you focus on key, important information
 - Prioritize: where will you spend your limited evaluation resources?
 - What do you really need to know??

Logic model and common types of evaluation



Types of evaluation

Needs/asset assessment:

What are the characteristics, needs, priorities of target population?

What are potential barriers/facilitators?

What is most appropriate to do?

Process evaluation:

How is program implemented?

Are activities delivered as intended? Fidelity of implementation?

Are participants being reached as intended?

What are participant reactions?

Outcome evaluation:

To what extent are desired changes occurring?
Goals met?

Who is benefiting/not benefiting? How?

What seems to work? Not work?

What are unintended outcomes?

Impact evaluation:

To what extent can changes be attributed to the program?

What are the net effects?

What are final consequences?

Is program worth resources it costs?

Foci of Today's Presentation

- 21st Century Skills – What are they? Why so much interest? Relevance to STEM education?
 - Relevant background ideas about the domain and the implications for instruction and assessment
- Examples drawn from Engineering Education
 - Assessment of conceptual understanding
- Assessment in the Context of Program Evaluation
- **Some Implications for Assessment and Evaluation in Engineering Education**

Some Possible Implications

- Important outcomes of engineering education align well with general goals for education & prediction of workplace success
 - *Deeper learning*
 - *21st Century Competencies*
- We develop these competencies through well designed disciplinary instruction and programs
 - Transfer is challenging and is not universal
 - Instructional design can foster deeper learning
 - Assessment is a key part of instructional design

Some Possible Implications

- Assessment of critical competencies is challenging and requires principled approaches to assessment design and validation
 - Cognitive, interpersonal, intrapersonal
 - Formative and summative assessment purposes
 - Concept inventories can be valuable in this effort
- Evaluation requires us to specify our *Theory of Action* and we can do so using *Logic Models*
 - Helps clarify intents, goals, and actions
 - Framework for formative & summative evaluation

gracias por su atención