Measuring Science Learning

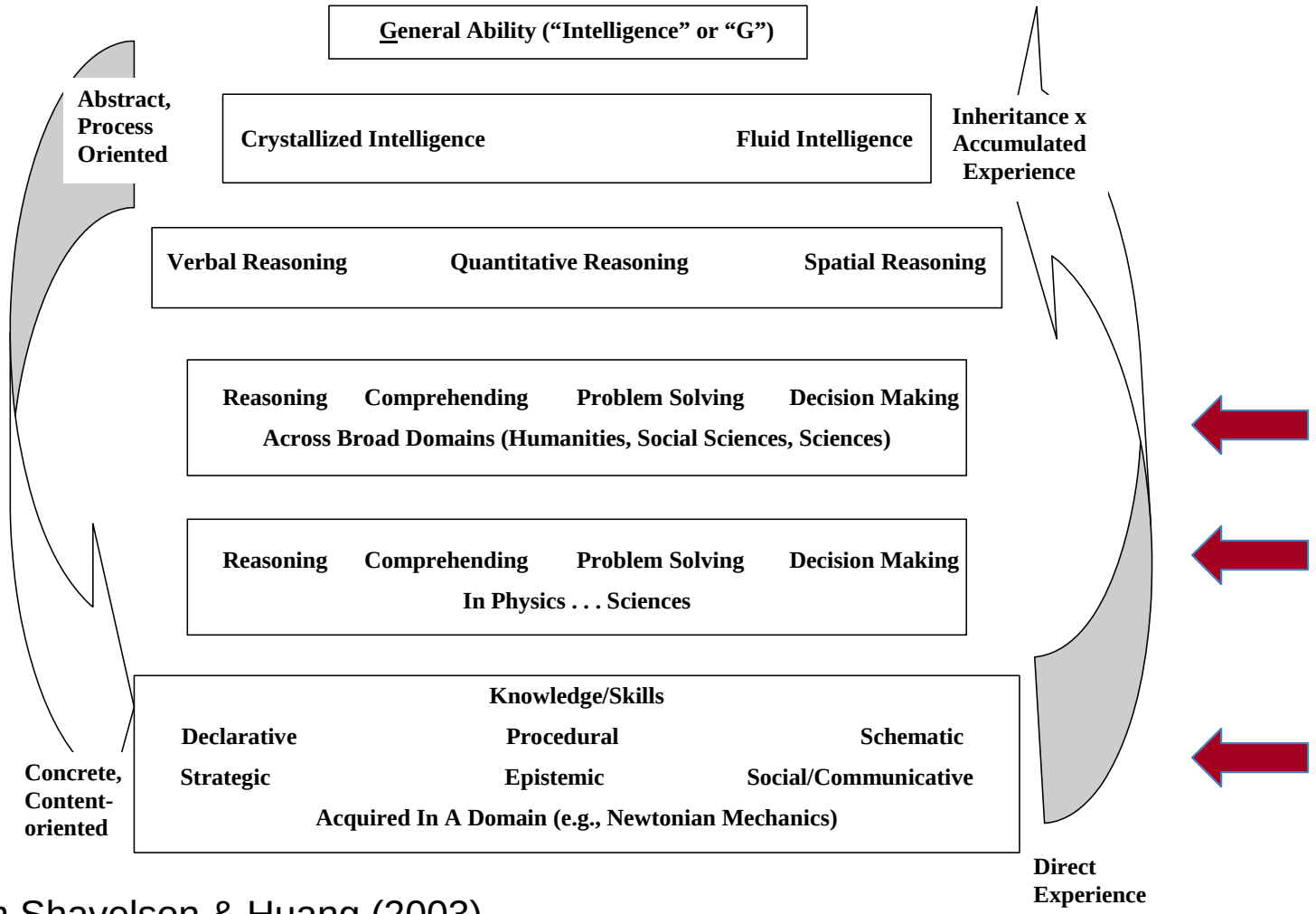
Rich Shavelson
Stanford University

Invited Presentation
Carl Wieman Science Education Initiative
March 25, 2010

Setting Learning Goals: What Do You Want To Emphasize

- Knowledge and skills?
- Critical thinking, analytic reasoning and problem solving?
- Habits of mind and epistemology?
- Individual and social understandings of the roles and responsibilities?

Psychological Framework Locating Science Achievement



Adapted from Shavelson & Huang (2003)

What Does It Mean To Achieve In Science?

- *Declarative knowledge*: knowing that--facts and concepts in the domain
- *Procedural knowledge*: routine procedures and some aspects of problem solving
- *"Schematic" (analytic) knowledge*: conceptual models of how the natural world works
- *Strategic ("transfer") knowledge*: knowing when, where and how knowledge applies
- *"Epistemic" knowledge:* knowing how we know—knowing how scientific knowledge is built and justified
- *Communication & social skills*: ability to communicate ideas clearly and concisely in the genre of science, team work

Knowledge Type Verbs

- Declarative—term or concept

  Define or describe, List/name characteristics, Relate to other, Exemplify, Classify

- Procedural:
  - Known procedures or steps:

    Recognize, Select, Execute
  - Data/Design

    Collect, Measure, Record, Represent, Interpret, Control

- Schematic—natural phenomenon

  Explain, Predict, Infer, Apply model, Pose (Q), Synthesize, Integrate

- Strategic—novel situation

  Conceptualize, Pose questions, Apply related knowledge/reasoning

Bloom's Taxonomy of the Cognitive Domain (Levels of Learning)

1. Factual Knowledge: remember and recall factual information
   Define, List, State, Label, Name, Describe
2. Comprehension: demonstrate understanding of ideas, concepts
   Describe, Explain, Summarize, Interpret, Illustrate
3. Application: apply comprehension to unfamiliar situations
   Apply, Demonstrate, Use, Compute, Solve, Predict, Construct, Modify
4. Analysis: break down concepts into parts
   Compare, Contrast, Categorize, Distinguish, Identify, Infer
5. Synthesis: transform, combine ideas to create something new
   Develop, Create, Propose, Formulate, Design, Invent
6. Evaluation: think critically about and defend a position
   Judge, Appraise, Recommend, Justify, Defend, Criticize, Evaluate

Higher level: Require deeper conceptual understanding
source of particular concern, but lower still matter.

Source: Carl Wieman @ Harvard Physics Dept.

Comparison of Verb Sets

Knowledge Type Verbs

Declarative—term or concept

Define or describe, List/name characteristics, Relate to other, Exemplify, Classify

Procedural:

- Known procedures or steps:

  Recognize, Select, Execute

- Data/Design

  Collect, Measure, Record, Represent, Interpret, Control

Schematic—natural phenomenon

Explain, Predict, Infer, Apply model, Pose (Q), Synthesize, Integrate

Strategic—novel situation

Conceptualize, Pose questions, Apply related knowledge/reasoning

Bloom Type Verbs

Factual Knowledge: remember and recall factual information
Define, List, State, Label, Name, Describe

Comprehension: demonstrate understanding of ideas, concepts
Describe, Explain, Summarize, Interpret, Illustrate

Application: apply comprehension to unfamiliar situations
Apply, Demonstrate, Use, Compute, Solve, Predict, Construct, Modify

Analysis: break down concepts into parts
Compare, Contrast, Categorize, Distinguish, Identify, Infer

Synthesis: transform, combine ideas to create something new
Develop, Create, Propose, Formulate, Design, Invent

Evaluation: think critically about and defend a position
Judge, Appraise, Recommend, Justify, Defend, Criticize, Evaluate

Assessing Declarative Knowledge Multiple-Choice: *TIMSS Pop.* 2

Air is made up of many gases. Which gas is found in the greatest amount?

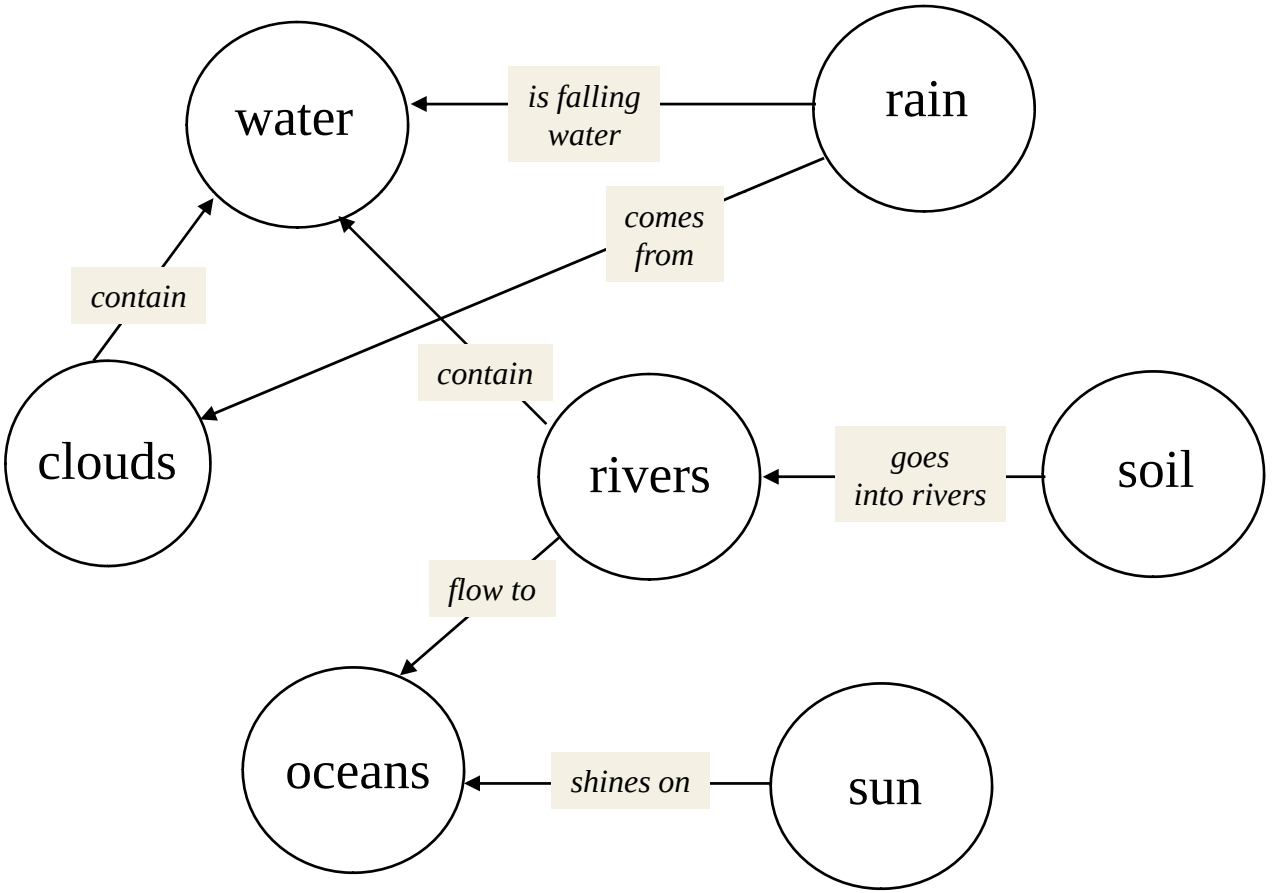
A. Nitrogen

B. Oxygen

C. Carbon Dioxide

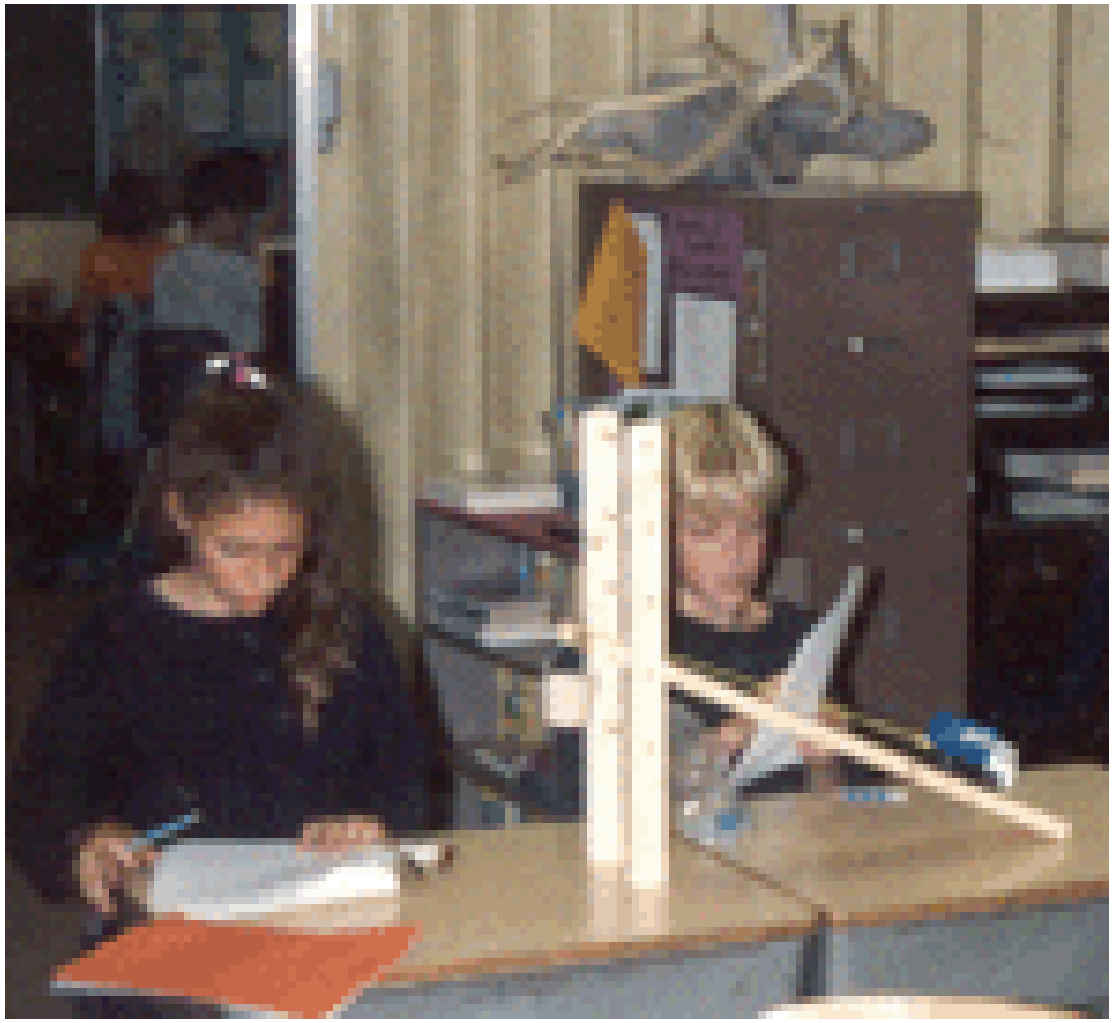
D. Hydrogen

Assessing Declarative Knowledge Structure: 11-Year-Old's Concept Map



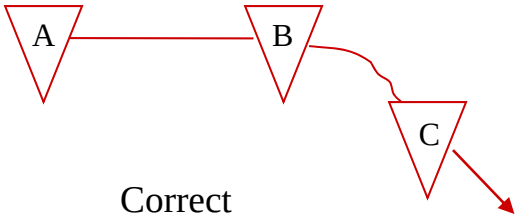
Source: White & Gunstone: *Probing Understanding* (1992, p. 16).

Assessing Procedural Knowledge: Incline Plane

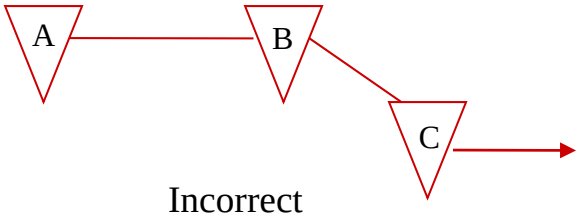


Assessing Schematic Knowledge: *Mental Models in Physics*

- (A) A rocket is moving along sideways in deep space, with its engine off, from point A to point B. It is not near any planets or other outside forces. Its engine is fired at point B and left on for 2 sec while the rocket travels from point B to point C. Draw in the shape of the path from B to C. (Show your best guess for this problem even if you are unsure of the answer.)
- (B) Show the path from C after the engine is turned off on the same drawing.



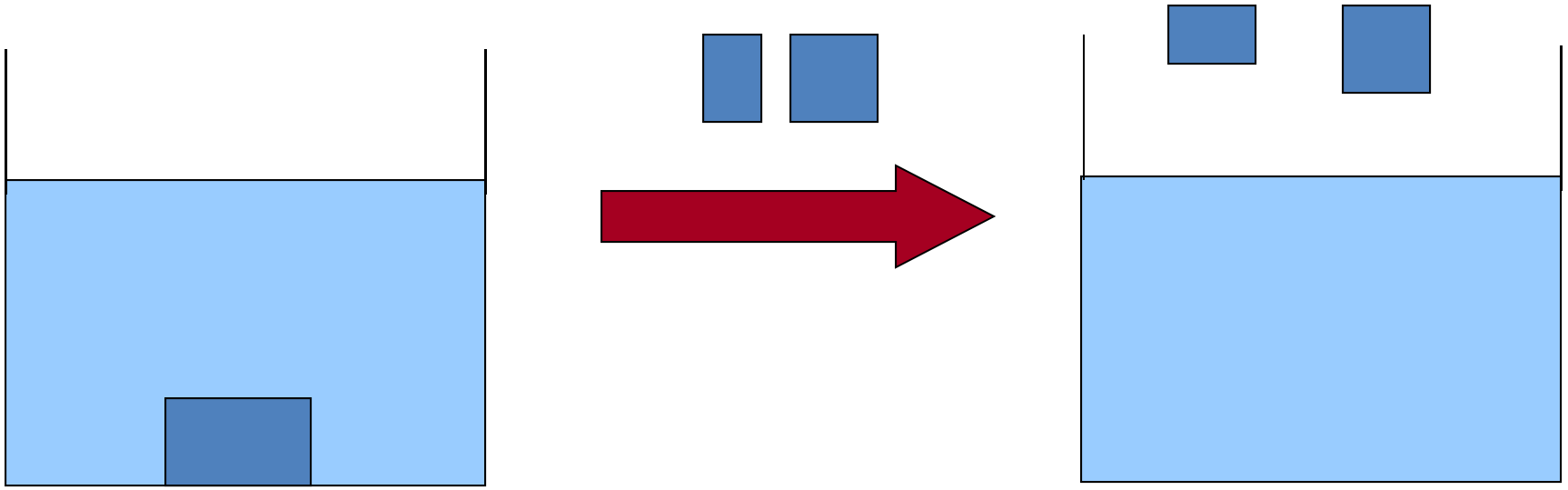
Correct



Incorrect

Source: Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66-71.

Assessing Schematic Knowledge: Predict-Observe-Explain



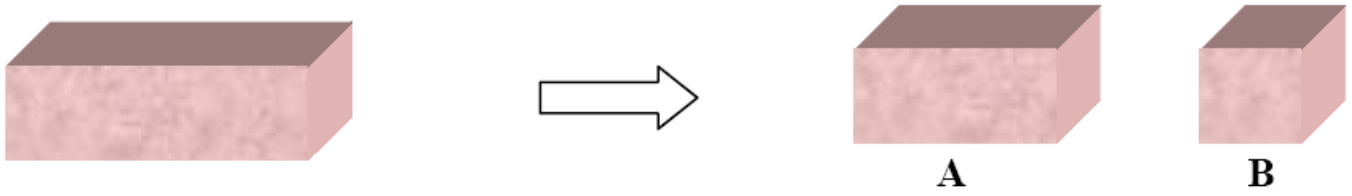
A soap bar sinks

Cut it into two unequal parts (1/3, 2/3)

What will happen to each?

Multiple-Choice Version of POE

Rich cut a wood block into two unequal parts. Part A is the 2/3 of the original one and part B is 1/3 of the original one. Which of the following relationships is correct?



A. The volume of A is the same as the volume of B.

B. The mass of A is the same as the mass of B.

C. The density of A is the same as the density of B.

D. The weight of A is the same as the weight of B.

TIMSS-R & Delaware Student Testing Program

Distribution of Science Test Items across Knowledge Types (Percent)

	Knowledge Type		
TEST	Declarative	Procedural	Problem-Solving
TIMSS-R	57.5	19.2	21.9
DSTP	56.0	20.0	24.0

Distribution of Science Test Items across Knowledge Types and Item Format (Percent)

		Knowledge Type		
Test	Format	Declarative	Procedural	Problem-Solving
TIMSS-R	Multiple-Choice	49.0	15.0	8.0
	Open-Ended	9.0	5.0	14.0
DSTP	Multiple-Choice	42.0	12.0	10.0
	Open-Ended	14.0	8.0	14.0

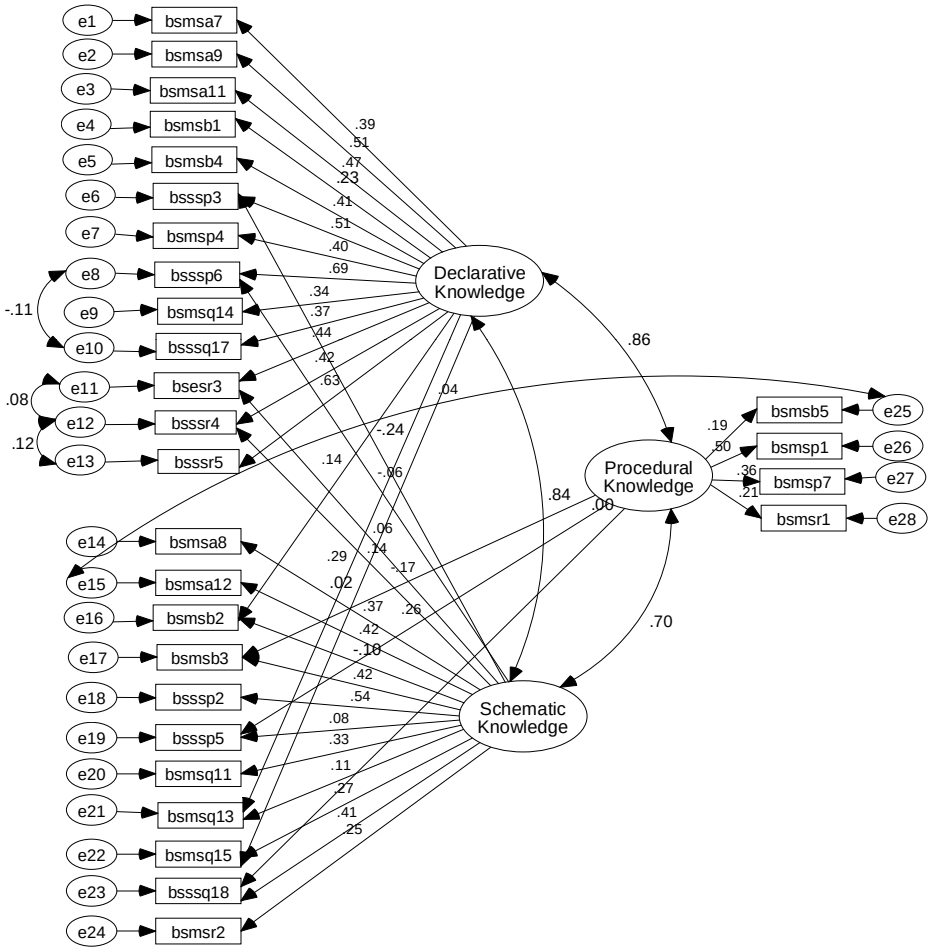
Cognitive Analysis: Link Between Logical And Cognitive Analysis

Based on the knowledge-type construct of science achievement, we expected participants' use of knowledge inferred from the protocols (cognitive analysis) to be congruent with the knowledge-types demanded by test items (logical analysis)

Type of knowledge used	Pre-classified knowledge-type			
	Declarative	Procedural	Schematic	Strategic
	(*n*=9)	(*n*=10)	(*n*=9)	(*n*=2)
Declarative	48*	8	11	0
Procedural	0	54	7	9
Schematic	9	16	41	0
Strategic	2	12	2	10

*Number of responses; Chi-square = 208.12, p<.001

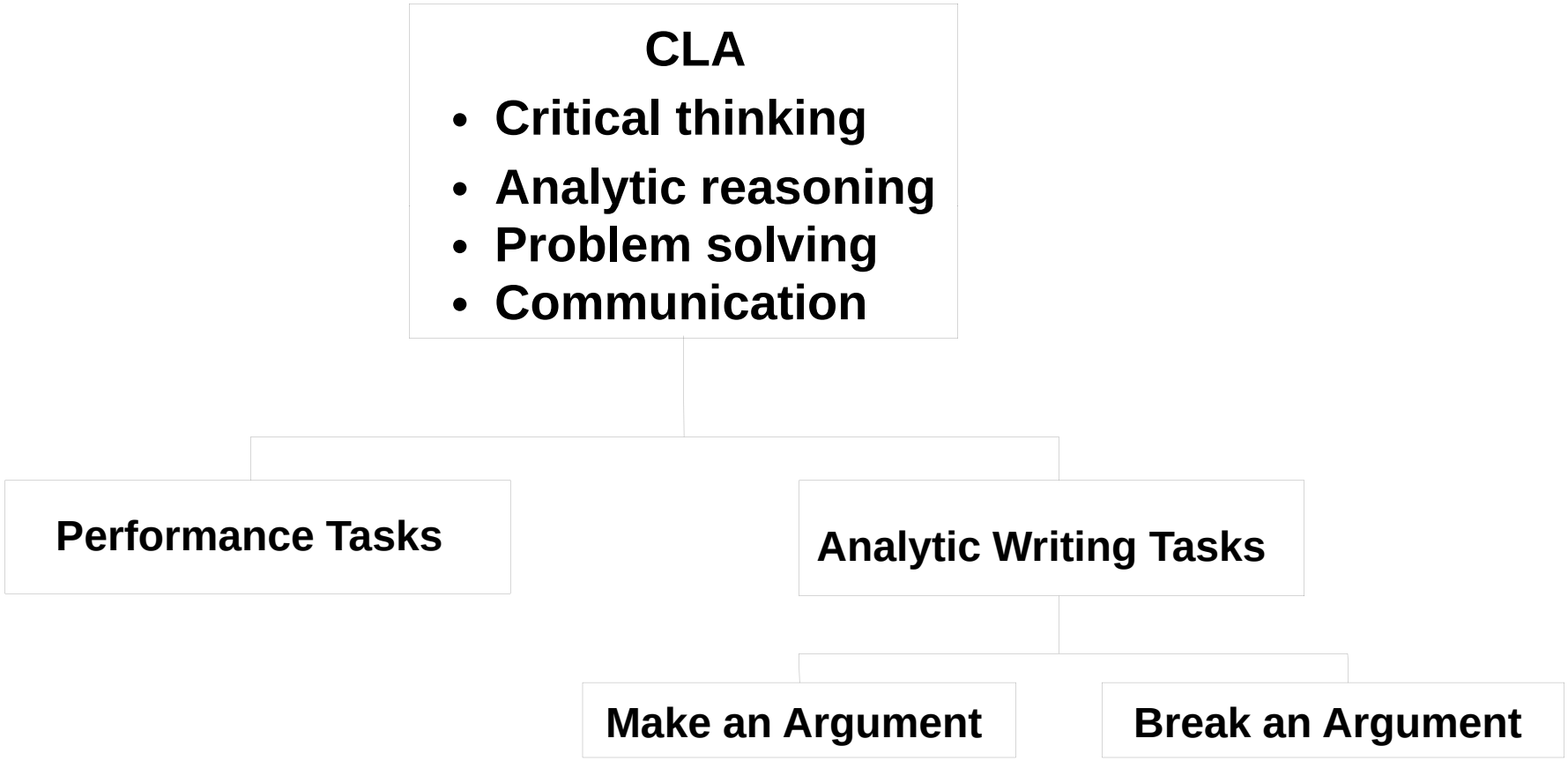
Statistical Modeling (Cont'd)



Source: Li (2001)

- A good statistical fit:
  - $\chi 2$=357.47, df=333, P=.17
  - CFI=.999
- Knowledge-type items clustered together as predicted.
- Declarative, procedural, and schematic knowledge factors highly correlated.
- Comparison with alternative models (e.g., one general factor, subject-matter factors) favored the knowledge-factor model.

The Collegiate Learning Assessment



What Is a [cla] Performance Task?

[collegiatelearningassessment]

Task Format

- Real-world problem
- Holistic, complex problem
- Information that may:
  - Be relevant or irrelevant to problem
  - Be reliable or unreliable
  - Lead to judgmental errors (e.g., correlation not causality, representativeness)

Response Format

- Make recommendation or decision, reach a conclusion, or solve a problem
- Minimally structured to support line of argument
- Written and not selected
- Requires evaluation of possible alternatives

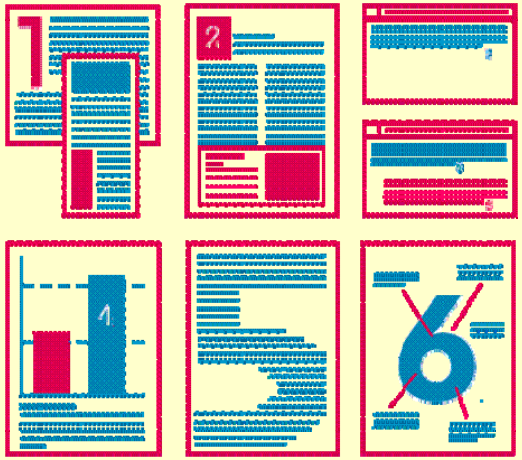
Scoring

- Analytic reasoning and evaluation
  - Identifies strengths and weaknesses of alternative arguments
  - Accurately judges quality of evidence avoiding unreliable, invalid, and erroneous information
- Problem solving
  - Provides decision and solid rationale based on credible evidence
  - Acknowledges uncertainty and need for further information
- Writing effectiveness
  - Organizes “advice” in logically cohesive and easy-to-follow way
  - Provides valid and comprehensive details supporting each argument and information source on which based
- Writing mechanics
  - Writes well constructed complex sentences
  - Shows outstanding control of grammar conventions
  - Demonstrates adept use of vocabulary

CLA Performance Task: "DynaTech" (90 Minutes)

You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech's sales force, recommended that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235. You are provided with the following documentation:

1: Newspaper articles about the accident
2: Federal Accident Report on in-flight breakups in single engine planes
3: Pat's e-mail to you & Sally's e-mail to Pat
4: Charts on SwiftAir's performance characteristics
5: Amateur Pilot article comparing SwiftAir 235 to similar planes
6: Pictures and description of SwiftAir Models 180 and 235



Please prepare a memo that addresses several questions, including what data support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups, what other factors might have contributed to the accident and should be taken into account, and your overall recommendation about whether or not DynaTech should purchase the plane.

CLA In-Basket Documents
"Crime" Performance Task

September 21, 2001

Jefferson Daily Press

Smart-Shop Robbery Suspect Caught
Drug-Related Crime on the Rise in Jefferson

Ann McNichols, Jefferson Township

On Monday police arrested a man suspected of robbing the Smart-Shop grocery store of $125. The arrest came less than six hours after J. Kim, the owner of the Smart-Shop store, reported the robbery. The suspect, David Koko, was found just a few blocks from the store and he put up no resistance when police arrested him. He was apparently high on drugs he had purchased with some of the money taken from the store. Mr. Kim told reporters that Koko came into the store just after it opened and demanded all the money from the cash register. He threatened the owner with a knife, and Mr. Kim gave him all the cash he had. The suspect fled, and Mr. Kim called the police.

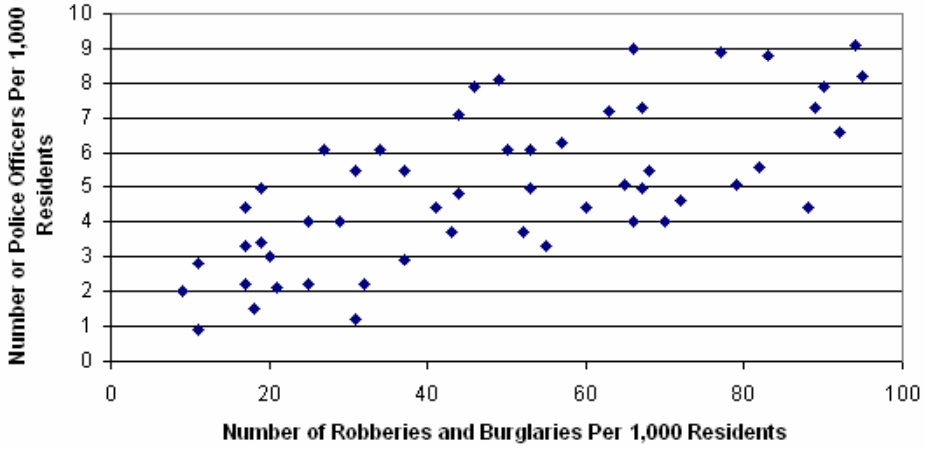
A few hours later police responded to a telephone complaint and found David Koko in an alley a few block from the store. The arresting officer said he appeared to be stoned and did not attempt to evade arrest. The officers found a syringe and other drug paraphernalia in Koko's pocket. He was charged with armed robbery and possession of drugs.

This is the fifteenth drug-related arrest in Jefferson this month, and the police are calling it an epidemic. Sergeant Hugh Morris said "Drugs are now the number one law enforcement problem in Jefferson. Half of our arrest involve drugs." Mayor Stone has called for more money to hire more police officers to reduce the growing crime rate in Jefferson. But the Council is divided on what to do. City Councilmen Slater and Colm called a press conference to demand that the rest of the council support an increase in the police budget. "If we put more cops on the street" they said, "we will show that criminals are not welcome in Jefferson." Mayoral candidate Dr. Jamie Hager called for a different approach. "More police won't make a difference, we need more drug treatment programs," Hager said. "The problem is not crime, per se, but crimes committed by drug users to feed their habit. Treat the drug use, and the crime will go away." The Council is slated to debate the proposed budget increase for police at its next meeting.

Crime Rate and Drug Use in Jefferson
By Zip Code

Zip Code	Percent of Population Using Drugs	Number of Crimes in 1999
11510	1	10
11511	3	20
11512	5	90
11520	8	50
11522	10	55

Crime Rates and Police Officers in Columbia's 53 Counties



CLA Make An Argument Writing

Directions: In 45 minutes, agree or disagree and explain the reasons for your position.

"In our time, specialists of all kinds are highly overrated. We need more generalists -- people who can provide broad perspectives."

CLA Break An Argument Writing

Directions: In 30 minutes, discuss how well-reasoned you find the argument.

A well-respected professional journal with a readership that includes elementary school principals recently published the results of a two-year study on childhood obesity. (Obese individuals are usually considered to be those who are 20 percent above their recommended weight for height and age.) This study sampled 50 schoolchildren, ages 5-11, from Smith Elementary School. A fast food restaurant opened near the school just before the study began. After two years, students who remained in the sample group were more likely to be overweight–relative to the national average. Based on this study, the principal of Jones Elementary School decided to confront her school's obesity problem by opposing any fast food restaurant openings near her school.

CLA Technology

Characteristic	Attributes
Open-ended Tasks	• Tap critical thinking, analytic reasoning, problem solving and written communication • Realistic work samples • Engaging task as suggested by alluring titles such as “brain boost,” “catfish,” “lakes to rivers”) • Applicable to different academic majors
Computer Technology	• Interactive internet platform • Paperless administration • Natural language processing software for scoring students written communication • Online rater scoring and calibration of performance tasks • Report institution’s (and subdivision’s) performance (and individual student performance confidentially to student)
Focus	• Institution or school/department/program within institutions • Not on individual student performance (although their performance is reported to them confidentially)
Sampling	• Samples students so that not all students perform all tasks • Samples tasks for random subsets of students • Creates scores at institution or subdivision/program level as desired (depending on sample sizes)
Reporting	• Controls for students’ ability so that “similarly situated” benchmark campuses can be compared • Provides value added estimates—from freshman to senior year or with measures on a sample of freshmen and seniors • Provides percentiles • Provides benchmark institutions

How We Know
Major-Specific Example: History

- Imagine a task asking history majors to explain (& justify) why this policy came about:

Discovery Day
October 21 Proclaimed A National Holiday By The President*

I Benjamin Harrison, President of the United States of America ... do herby appoint Friday, Oct. 21, 1892, the four hundredth anniversary of the discovery of America by Columbus, as a general holiday for the people of the United States. On that day let the people so far as possible cease from toil and devote themselves to such exercises as may best express honor to the discoverer and their appreciation of the great achievements of the four completed centuries of American life….

Let the national flag float over every school house in the country …. In the Churches and in the other places of assembly of the people….

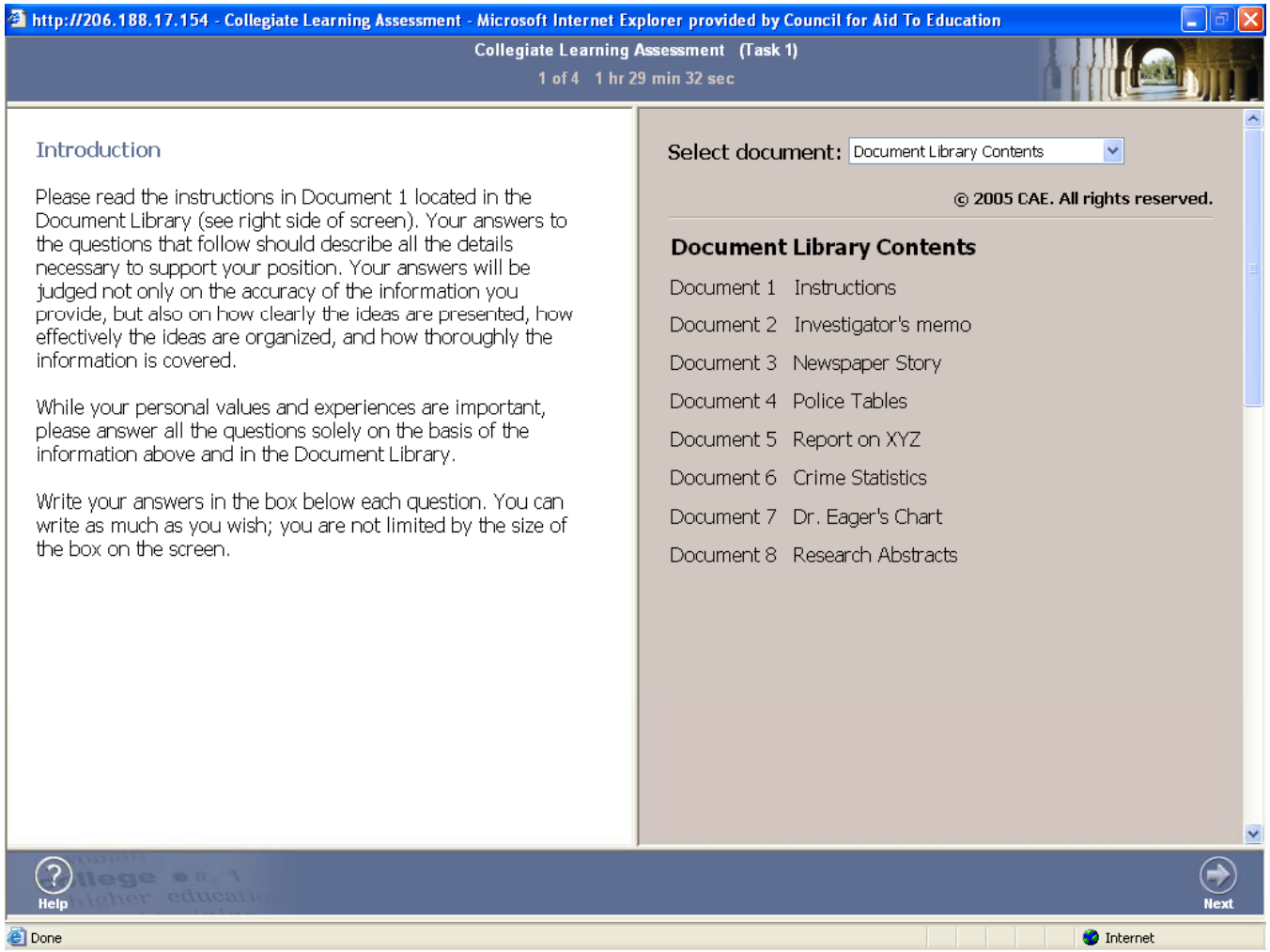
* New York Times, July 22, 1892, p. 8

- Accompanied by an basket of historical documents:
  - Statistics on immigration by country
  - Statistics on religious affiliation
  - News story on changing immigration
  - News story on developments in Russia

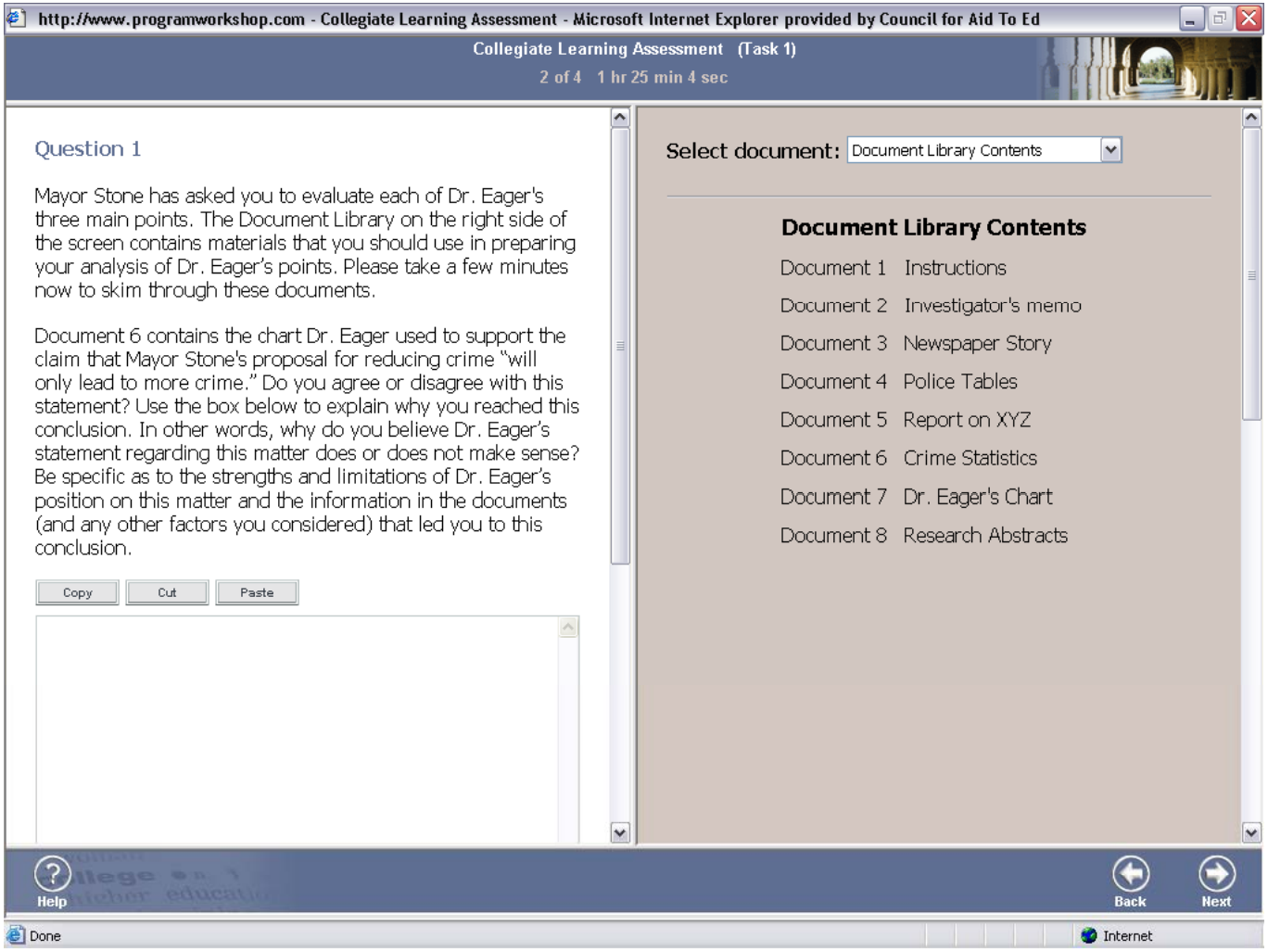
- Factors distinguishing *novices* (high school history students) and *experts* (history grad students):
  - Sourcing
  - Corroborating
  - Contextualizing

Thank You!

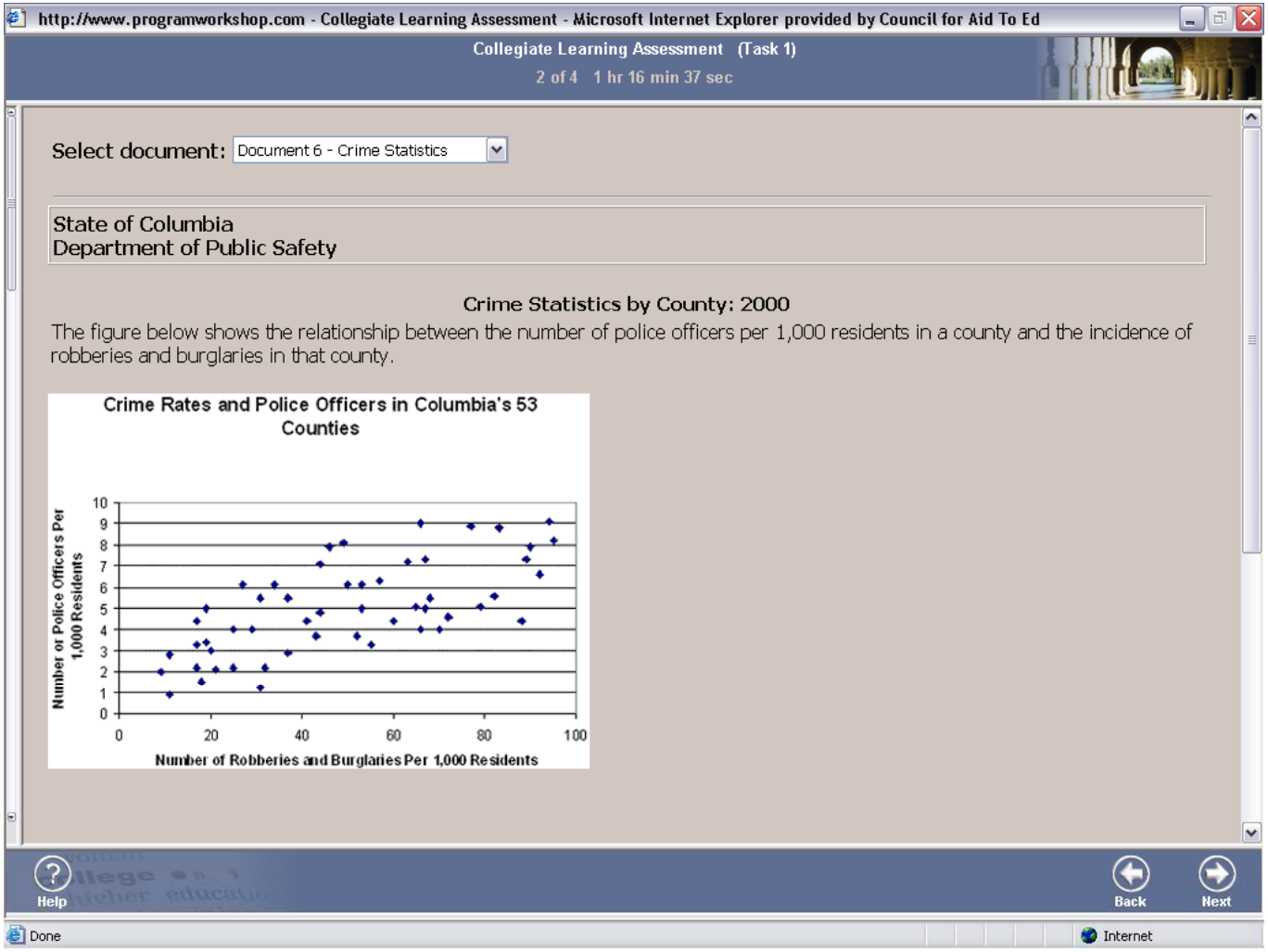
Internet Platform: Introduction To Task



Internet Platform: First Question



Internet Platform: Library Document



Reliability

- Grading
  - Inter-reader consistency
  - High hand/machine agreement rate
  - Continually monitor machine accuracy
- Test scores – split sample analyses – high correlations:
  - School means on a task
  - School difference (residual) scores within a class
  - School value-added scores across classes
- High correlations require reliable scores
- Characteristics of participating schools are similar to those in a national database

Reliability Evidence: Performance Tasks

Table 2. Estimated Variance Components in the Example $s \times t \times j$ Design.

Source	Variance Component	Estimate	%Total Variability
School (s)	σ_s^2	817.466	20.9
Task (t)	σ_t^2	0[a]	0
Judge (j)	σ_j^2	62.564	1.6
$s \times t$	σ_{st}^2	671.423	17.1
$s \times j$	σ_{sj}^2	62.178	1.6
$t \times j$	σ_{tj}^2	0[a]	0
$s \times t \times j, e$	$\sigma_{stj,e}^2$	2305.770	58.8

[a]Negative variance component set equal to zero.

Reliability (Generalizability) = 0.803