

# Assessing Student Learning Outcomes



Richard J. Shavelson  
Stanford Education Assessment Lab

Invited Presentation  
Texas Higher Education Coordinating Board  
2008 Texas Higher Education Leadership Conference  
*Closing the Gaps: Student Outcomes & Student Success*

Austin, Texas  
December 2, 2008

# Overview

---

- ❑ “Student learning” and other popular terms
- ❑ A brief journey through the history of student learning assessment
- ❑ Some current options for assessing learning
- ❑ The Collegiate Learning Assessment
- ❑ Concluding comments

# On Learning

---

□ **Learning: A relatively permanent change in behavior over time not due to maturation**

- *Longitudinal approach to assessing learning*: Measure a student's performance at two points in time:

Learning = Performance as a Senior – Performance as a Freshman

- *Cross-sectional (proxy) approach to assessing learning*: Measure freshmen's and seniors' average performance at one point in time:

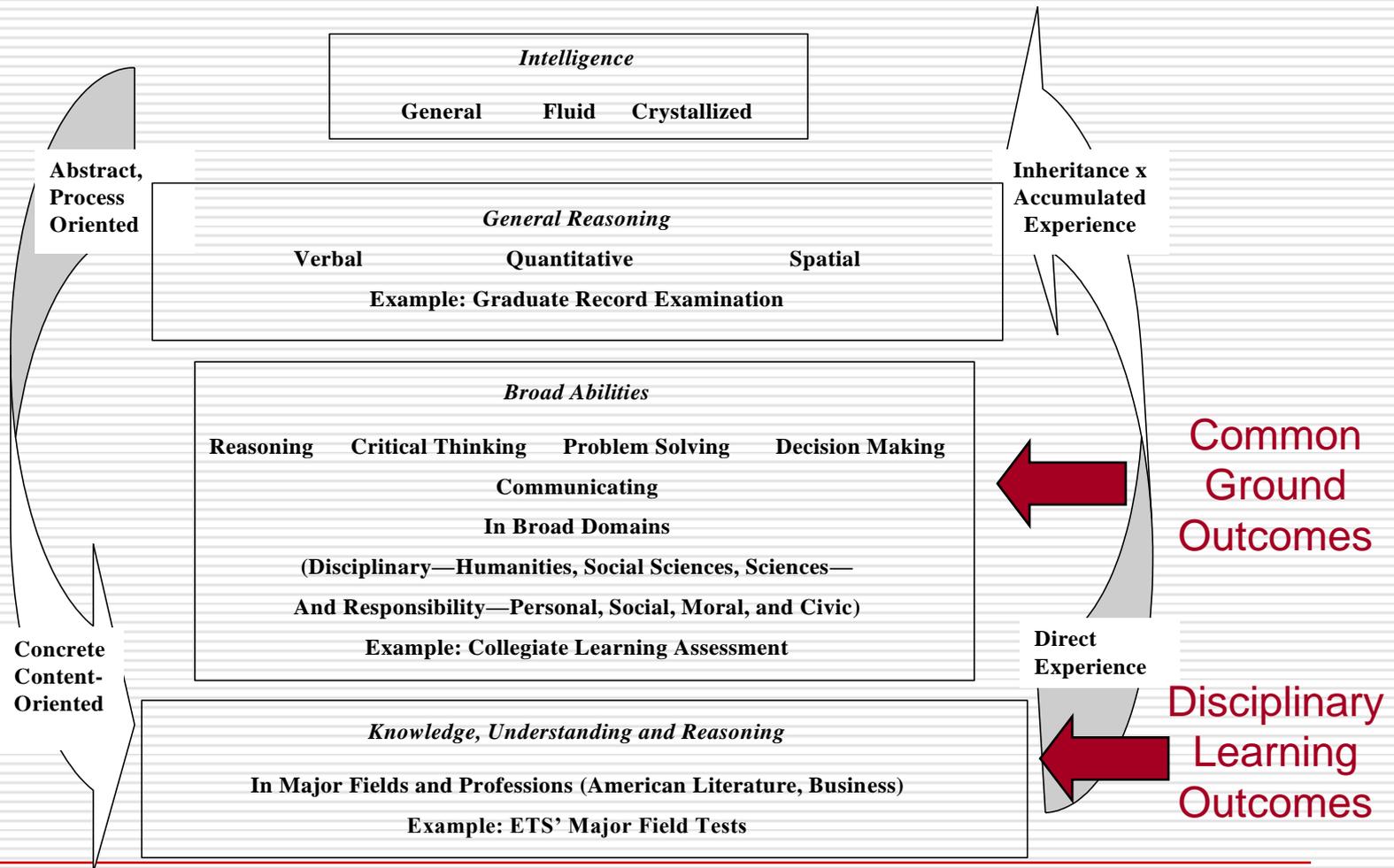
Learning = Average Performance of Seniors – Average Performance of Frosh *in 2008*

□ **Achievement: A snapshot of the level of a student's performance at one point in time**

□ **Value-added: Student learning (change) adjusted for what they would be expected to learn based on their abilities upon entering college (e.g., adjusted for entering SAT/ACT scores)**

---

# Framework For Learning Outcomes



# Brief History I: Carnegie Foundation for the Advancement of Teaching (1900 – 1933)

---

- Impetus: Concern with the quality of higher education
- Mechanism: New technology of objective testing
- Vision of learning
  - Accumulation of breadth and depth of content knowledge
  - “Relatively permanent and available equipment of the student”
  - “Must be so familiar and so sharply defined that it comes freely to mind when needed”
- Contrasting vision of learning (Progressive Education Association):
  - Reasoning and problem solving
  - Life tasks
  - Moral, self and social development

# Brief History I: Carnegie's Pennsylvania Study (1928 – 1932)

1928: Objective testing of *declarative* and *procedural knowledge (achievement)* of State's college seniors (70% or 4,580) and high-school seniors (75% or 26,500)

1928: Testing time was 12 hours; roughly half that later

1928: Number of test items = 3,200; roughly half that later

1930 & 1932: College *learning* of high-school seniors tested longitudinally both years

## IV. GENERAL SCIENCE, *Part II*

*Directions:* In the parenthesis after each word or phrase in the right-hand column, place the number of the word or phrase in the left-hand column of the same group which is associated with that word or phrase.

- |                                 |            |
|---------------------------------|------------|
| 14. 1. Unit of work             | Calorie    |
| 2. Unit of potential difference | Dyne       |
| 3. Unit of electrical current   | Erg        |
| 4. Unit of heat quantity        | H. P.      |
| 5. Unit of power                | Volt       |
| 6. Unit of force                | Ampere     |
| 7. Unit of pressure             | B. T. U.   |
|                                 | Atmosphere |
|                                 | Foot-pound |
|                                 | Watt       |

## V. FOREIGN LITERATURE, *Multiple Choice*

9. Sophocles' *Antigone* is a depiction of 1 the introduction of laws into a barbarous state, 2 the prevailing of sisterly love over citizenly duty, 3 idyllic peasant life, 4 the perils of opposing oneself to Zeus
10. Of Corneille's plays, 1 *Polyeucte*, 2 *Horace*, 3 *Cinna*, 4 *Le Cid*, shows least the influence of classical restraint

## VI. FINE ARTS, *True-False*

1. Greek architecture prior to contact with the Romans made no use of the dome
10. The slow movements of Beethoven's symphonies are somewhat inferior to the rest of those compositions

## VIII. MATHEMATICS

*Directions:* Each of the problems below is followed by several possible answers, only one of which is entirely correct. Calculate the answer for each problem; then select the printed answer which corresponds to yours and put its number in the parenthesis at the right.

5. If two sides of a triangle are equal, the opposite angles are  
(1) equal (2) complementary (3) unequal (4) right angles

Source: Leamed and Wood, 1938.

# Brief History II: Carnegie And The GRE

---

## □ Impetus:

- The AB degree had “ceased to draw the line between the fit and unfit”
- Something more than number of college (“Carnegie”) credits was needed for admission and graduate-student quality decisions

## □ Mechanism:

- Parlayed the Pennsylvania Study staff and tests into a graduate admissions test

# Brief History II: Rise & Changing Focus Of The GRE

---

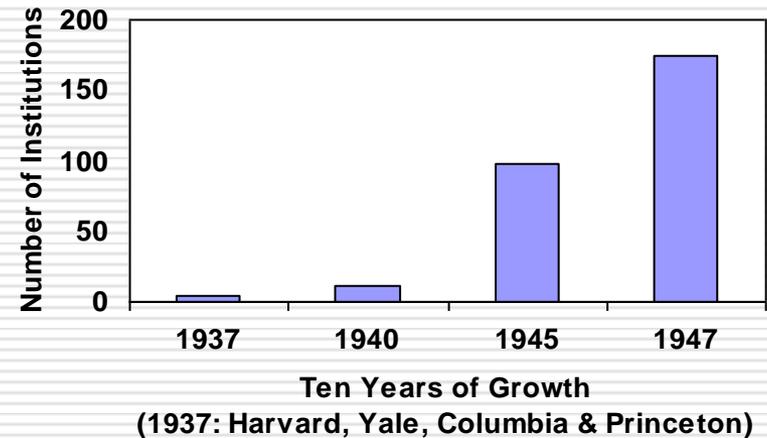
## □ Evolution of the GRE

1936: Achievement test of content knowledge in 8 subjects (e.g., physical sciences, mathematics, social studies, literature)

1946: Aptitude test of quantitative and verbal reasoning

2008: Aptitude test of quantitative and verbal reasoning and analytic writing

## □ Rapid adoption



# Brief History III: Rise Of The Test Providers

---

- ❑ End of WWII and GI Bill rapidly expands demand for higher education
- ❑ In response supply increases:
  - Number of colleges and universities mushroom
  - Number of test providers increases
    - ❑ ETS established in 1948
    - ❑ ACT established in 1959
- ❑ Carnegie creates ETS in response to:
  - Demand for testing after WWII
  - GRE testing program burden
  - Financial burdens created by its innovative faculty retirement program (for which The Foundation was created)
  - Need to preserve “meritocracy”

# Brief History III: Tests Of Undergraduates' Learning

---

- Tests in the Foundation's vision
  - ETS' Undergraduate Assessment Program
  - ACT's College Outcomes Measurement Project (COMP)
- Tests in the Progressive Education Association's vision
  - Jonathan R. Warren's study of "**free-response questions**" tapping communication, analytic thinking, synthesizing ability, social/cultural awareness (ETS)
  - **COMP** began as performance assessment of effective functioning in adult life in social institutions, in using science and technology, and in the arts (ACT)
  - **Tasks in Critical Thinking** were performance-based measures of critical thinking and communication skills
- **Each program either died out (Warren, Tasks) or morphed into multiple-choice format (COMP) due to time, cost, and technology constraints**

# Brief History IV: Era Of Accountability (1979 – Present)

---

- Spellings' Commission (2006)
  - *"Higher Education must change from a system primarily based on reputation to one based on performance"*
  - Urged *"creation of robust culture of accountability and transparency"*
  - Student achievement inextricably connected to *institutional success*
  - Achievement measured by *value added*
  - *Relative effectiveness* of campuses published in league tables
- Six professional associations agree (2006) higher-education needs to:
  - Improve learning
  - Increase accountability for educational outcomes
  - Extreme response from 2 associations:
    - Voluntary System of Accountability (VSA) to demonstrate accountability and stewardship to the public
    - College Portrait system of indicators
    - High Stakes: Publish comparative results of student achievement and value added

# Brief History IV: Academics' Reaction

---

- ❑ **One size does not fit all**—Commission failed to recognize diversity of higher education institutions
- ❑ **Outcomes vary by academic major**
- ❑ **Sole focus on “cognitive” outcomes** (declarative and procedural knowledge) too **limiting**—need individual and social responsibility outcomes
- ❑ **Intrudes on academic culture** where faculty responsible for curriculum, teaching and assessment
- ❑ **Higher education system too complex** for simple quantitative measurement
- ❑ **“Horse-race” comparisons** of colleges and universities at best **misleading** and at worse have **perverse** effect on teaching and learning

# Rapprochement: Common Ground For Higher Education

---

- ❑ Think critically
- ❑ Reason analytically
- ❑ Connect apparently disparate pieces of information
- ❑ Explore others' knowledge claims
- ❑ Justify own knowledge claims with evidence and examples
- ❑ Take another's perspective and act accordingly

# Some Current Options: Approaches To Measuring Student Learning

---

## **Ability (CAAP, MAPP)**

- Divide complex tasks and sample specific abilities
- Get estimates of stable underlying abilities
- Keep test items and answers secret
- Combine specific abilities actuarially
- Focus on respondent behavior
- Sample specific behavioral objectives

## **Criterion Sampling (CLA)**

- Sample criterion situations
- Reflect changes in what individual has learned
- Show how to improve on performances assessed (feedback)
- Focus on competencies in *clusters* of life outcomes
- Involve operant as well as respondent behavior
- Sample operant thought patterns

# Some Current Assessment Options I: *The Carnegie Tradition - Ability*

---

## Collegiate Assessment of Academic Proficiency (CAAP—ACT's newer version of COMP)

### Science Example:

A scientist investigated the factors that affect seed mass in the plant species *Desnodium poniculatum*. Some results ... are summarized in the two tables below.

The data suggest that subjecting plants to which of the following conditions would result in the greatest seed masses?

- A. 8 hours of light, adequate water supply and 23 C
- B. 8 hours of light, decreased water supply and 23 C
- C. 14 hours of light, adequate water supply and 23 C
- D. 14 hours of light, decreased water supply and 29 C

## Measure of Academic Proficiency and Progress (MAPP—ETS' renovated Academic Profile)

### Writing Example:

Being a female jockey, she was often interviewed

Rewrite, beginning with She was often interviewed

The next words will be

- (A) on account of she was
- (B) by her being
- (C) because she was
- (D) being as she was

### Math Example:

From 6:00 a.m. to 7:00 a.m. one morning, the temperature rose 7 F. From 7:00 a.m. to 8:00 a.m., the temperature rose 2 F, and from 8:00 a.m. to 9:00 a.m., the temperature rose 3 F. If the temperature was 55 F at 9:00 a.m., what was the temperature at 6:00 a.m.?

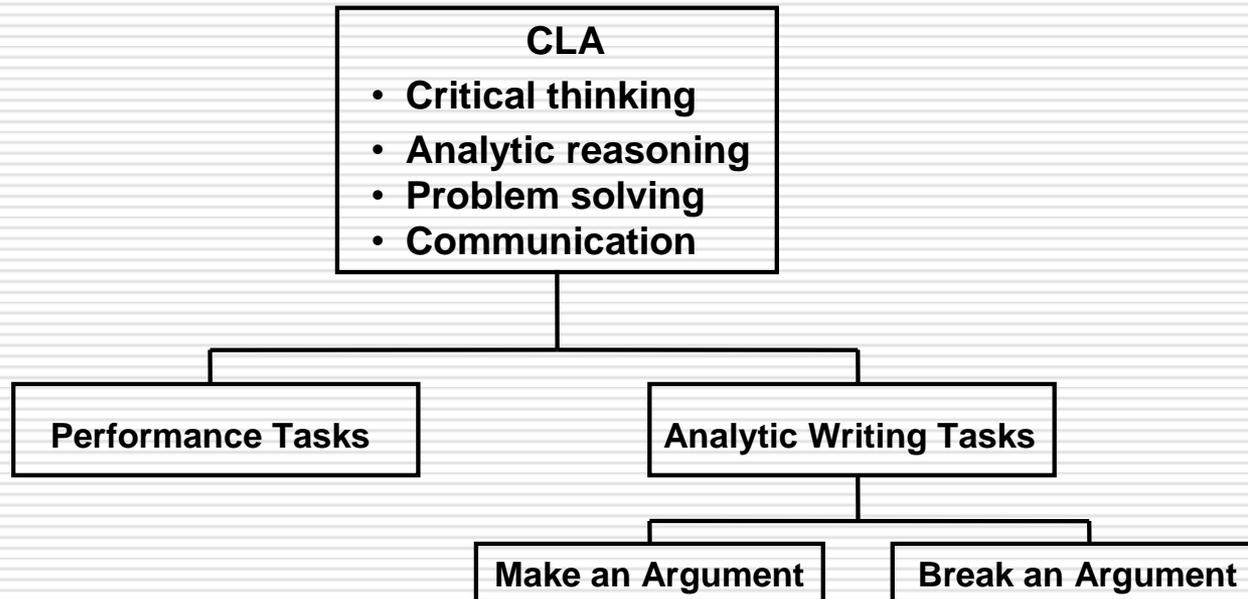
- (A) 43 F
- (B) 53 F
- (C) 57 F
- (D) 67 F

# Some Current Assessment Options II:

*Progressive Education Era Tradition - Criterion Sampling*

---

## The Collegiate Learning Assessment (CLA)

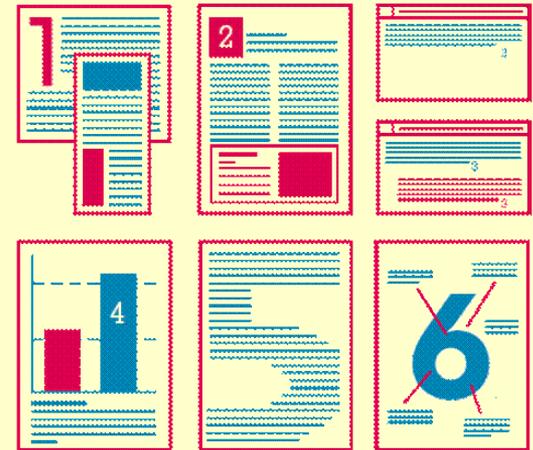


# CLA Performance Task: Example I— “DynaTech” (90 Minutes)

---

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 Performance Task: Example II— “Crime” Performance Task In-Basket Docs

September 21, 2001

## Jefferson Daily Press

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

Ann McNickle, Jefferson Daily Press

On Monday police arrested a man suspected of robbing the Smart-Shop grocery store of \$125. The arrest came less than an hour after J. Kim, the owner of the Smart-Shop store, reported the robbery. The suspect David Keke, 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 Keke 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.

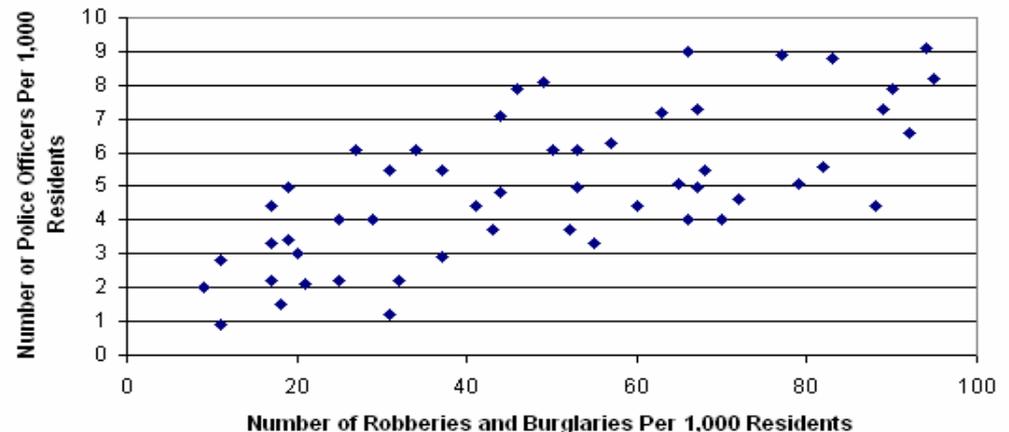
A few days later police responded to a telephone complaint and found David Keke manally a few blocks from the store. The arresting officer said he appeared to be stoned and did not attempt to evade arrest. The officers found a syringe and other drug paraphernalia in Keke's pocket. He was charged with 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 "Drug are now the number one law enforcement problem in Jefferson. Half of our arrests 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 Eager called for a different approach. "More police won't make a difference, we need more drug treatment programs," Eager 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	<ul style="list-style-type: none"> <li>• Tap critical thinking, analytic reasoning, problem solving and written communication</li> <li>• Realistic work samples</li> <li>• Engaging task as suggested by alluring titles such as “brain boost,” “catfish,” “lakes to rivers”)</li> <li>• Applicable to different academic majors</li> </ul>
Computer Technology	<ul style="list-style-type: none"> <li>• Interactive internet platform</li> <li>• Paperless administration</li> <li>• Natural language processing software for scoring students written communication</li> <li>• Online rater scoring and calibration of performance tasks</li> <li>• Report institution’s (and subdivision’s) performance (and individual student performance confidentially to student)</li> </ul>
Focus	<ul style="list-style-type: none"> <li>• Institution or school/department/program within institutions</li> <li>• Not on individual student performance (although their performance is reported to them confidentially)</li> </ul>
Sampling	<ul style="list-style-type: none"> <li>• Samples students so that not all students perform all tasks</li> <li>• Samples tasks for random subsets of students</li> <li>• Creates scores at institution or subdivision/program level as desired (depending on sample sizes)</li> </ul>
Reporting	<ul style="list-style-type: none"> <li>• Controls for students’ ability so that “similarly situated” benchmark campuses can be compared</li> <li>• Provides value added estimates—from freshman to senior year or with measures on a sample of freshmen and seniors</li> <li>• Provides percentiles</li> <li>• Provides benchmark institutions</li> </ul>

# CLA Report: Summary

University College contributes more to the learning gains made by students than 96 percent of the 176 four-year undergraduate institutions participating in the 2007–2008 CLA. University College performed Well Above Expected.

	Freshmen		Seniors		Value-Added Estimate	
	<i>Percentile Rank</i>	<i>Performance Level</i>	<i>Percentile Rank</i>	<i>Performance Level</i>	<i>Percentile Rank</i>	<i>Performance Level</i>
<i>Total CLA Score</i>	39	At	92	Well Above	96	Well Above
<i>Performance Task</i>	35	At	86	Above	94	Well Above
<i>Analytic Writing Task</i>	51	At	89	Above	86	Above
<i>Make-an-Argument</i>	71	Above	70	Above	50	At
<i>Critique-an-Argument</i>	27	Below	97	Well Above	98	Well Above

# Some Current Options III: GRE?

## GRE-Quantitative Reasoning

You are to write your answers in the boxes below the problems. Several practice problems are given below with the first one correctly worked. Practice for speed on the others. This practice may help your score.

### Practice Problems:

4	7	12	84	7	34	17	45	31	80
9	6	5	54	38	81	50	41	52	78
1	15	67	72	80	51	74	89	19	15
$\frac{1}{4}$	<input type="text"/>								

Your score on this test will be the number of problems that are added correctly. Work as rapidly as you can without sacrificing accuracy.

In this test you will be asked to solve some problems in arithmetic. Work each problem and put an x on the number in front of the answer that you choose.

Example:

How many candy mints can you buy for 50 cents at the rate of 2 for 5 cents?

- 1 - 10
- 2 - 20
- 3 - 25
- 4 - 100
- 5 - 125

The correct answer to this problem is 20. Therefore, you should have marked an x through the number 2 to indicate the correct answer.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

## CLA-Like Version Quantitative-Reasoning

“One of the drugs in the Coronary Drug Project was nicotinic acid. Suppose the results on nicotinic acid were as reproduced below. Something looks wrong. What, and why?”

Group	Nicotinic Acid		Placebo	
	Number	Deaths	Number	Deaths
Adherers	558	13%	1813	15%
Non-adherers	487	26%	882	28%
Total	1045	19%	2695	19%

Source: Friedman *Statistics*

# Concluding Comments

---

- As the Spellings Commission pointed out, a campus' success is measured by student learning
  - We have a long tradition of measuring student learning in higher education; often lessons are re-learned when this rich history is forgotten
  - A variety of student learning assessments are currently available, drawn largely from the Carnegie tradition of ability measurement:
    - CAAP
    - MAAP
    - GRE
  - The CLA
    - Stands in contrast to other measure of undergraduates' learning, coming from a progressive tradition of reasoning, problem solving, critical thinking and communicating
    - Focuses on improving teaching and learning: "We encourage institutions to (1) communicate results across campus, (2) link student-level CLA results with other data sources, (3) pursue in-depth sampling, and (4) participate in CLA in the classroom"
-

---

Thank You!