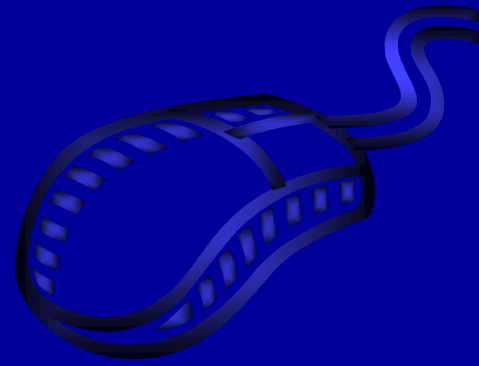# And Now For Something Completely Different…

Selected Novel Test Item Formats for Computerized Assessment

April L. Zenisky

University of Massachusetts at Amherst

February 2005

Testing and computer technology

- Use of computer technology in many aspects of assessment is often a matter of course
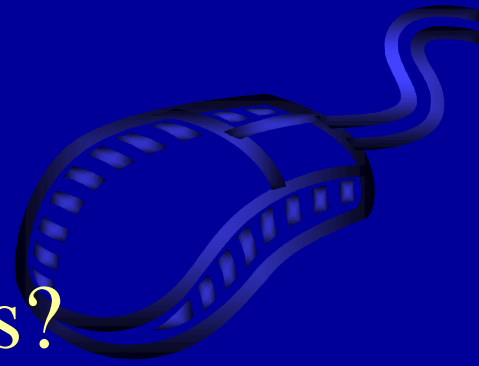
Development

Analysis

Reporting

Scoring

and increasingly, Administration

Technology and CBT

- CBT as a medium for <u>giving</u> tests is maturing

  – Away from static linear forms

  – Increased interest in taking advantage of graphical, video, and audio capabilities

- Reconsideration of what examinee responses could look like (Bennett, 1998)

  – Goal: enhancing measurement information

Novel item types

- Why develop or use new item types?
  - Potential for greater alignment of test items with construct(s) of interest
    - For example, NBME's Step 3 Computer-Based Case Simulations (CCS); also, the AICPA's new CPA Exam
  - Increased interest among test developers and examinees in more "active" or "engaging" assessment methods
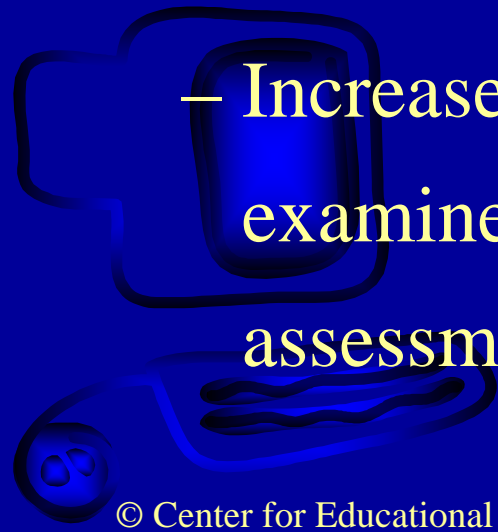
Focus for today

- A framework for thinking about items and item types in computer-based testing

- Lots of examples

- Issues and research needs

Purposes

- Introduce the possibilities that are out there for computerized testing

- Think about tests and test items in terms of information value

  - What do different item types bring to measuring a construct?

- Talk about the state of research and identify needs

Caveat to the Presentation

- No skeptics allowed (until later)
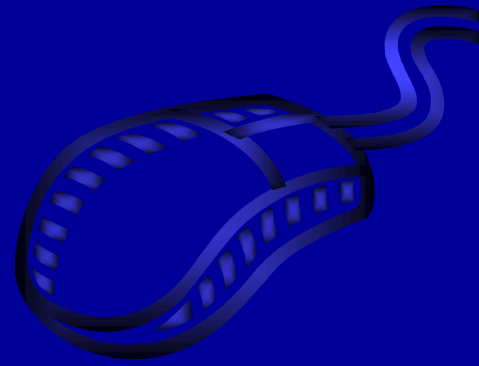
What is a test item?
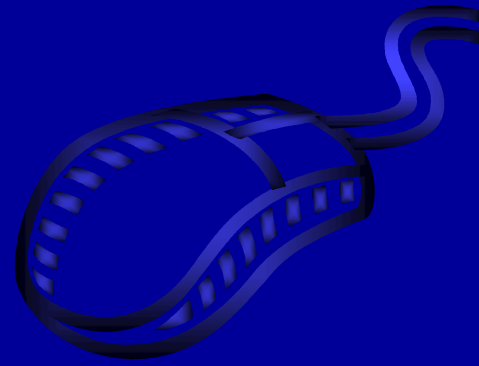
- From Osterlind (1990):

> "…a unit of measurement with a stimulus and prescriptive form for answering; (…) it is intended to yield a response from an examinee from which performance in some psychological construct may be inferred"

What is a test item?

- The first part of a test item is the *stem*
  - This is where a question is asked or a task to be completed is laid out
 - Must be explicitly clear to examinees what they need to do

What is a test item?

- The second part of a test item is the *response*, where students answer as directed in the stem

- Can take many forms
  - Selected or constructed response
 - More constrained to less constrained

Response actions

- Parshall, Davey, & Pashley (2000)

  - What should examinees know and be able to do relative to a specified construct?

  - What kind of information should be gathered from examinees?

  - In what ways can that information be represented?

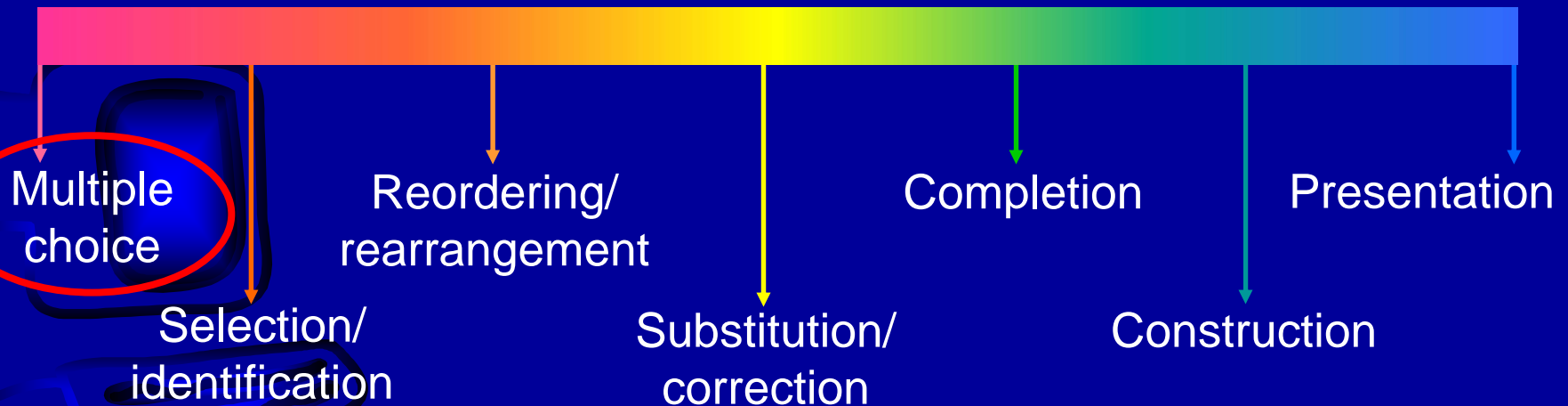 - How should it be entered into the computer?
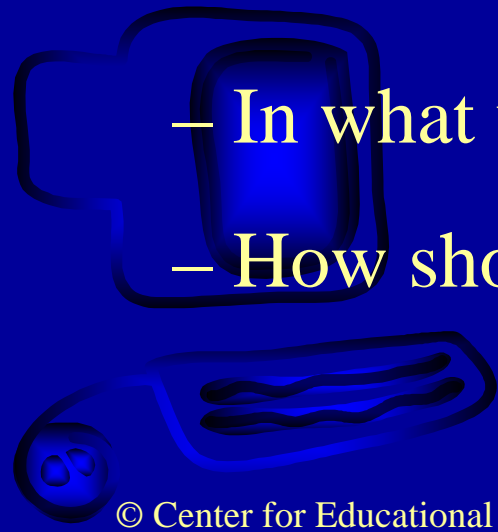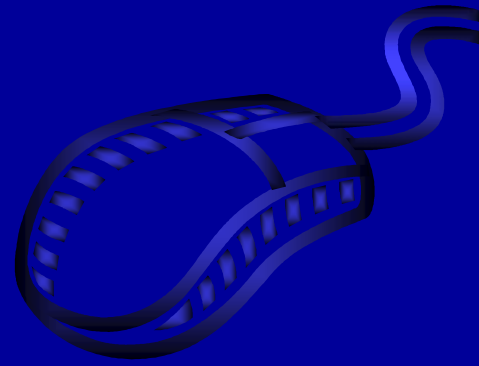
Response actions (cont.)
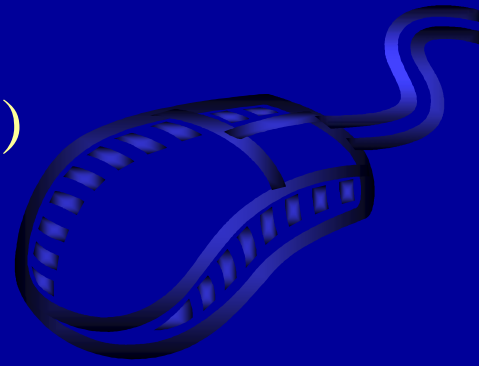
- In CBT, these commonly include:
  - Typing response via keyboard
  - Using mouse to drag-and-drop onscreen items
  - Drawing/illustrating concepts with onscreen tools
  - Clicking on screen with mouse

- Perhaps someday, with more research:
  - Touch screens, light pens, joysticks, trackballs, speech-recognition software
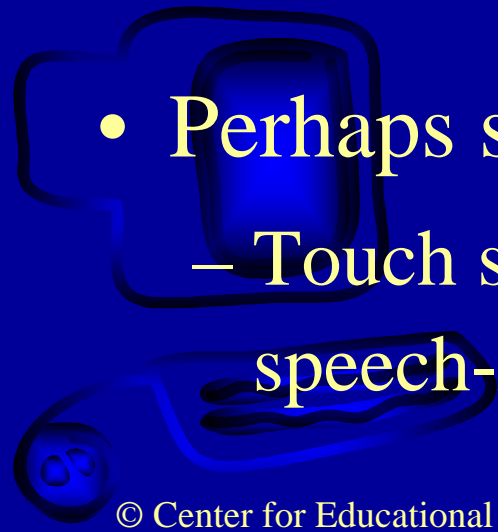
Conceptualizing Item Types

- Continuum from Bennett, Ward, Rock, and LaHart (1990)

  – Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)
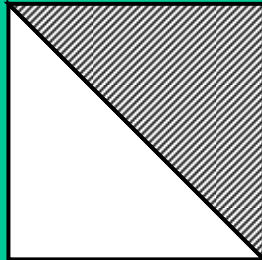
Multiple choice

Selection/ identification

Reordering/ rearrangement

Substitution/ correction

Completion

Construction

Presentation

Multiple-choice

- Choose an answer from a small set of response options

- Perception that MC items only assess recall
  - The research is equivocal (Martinez, 1999)
    - Is it a limitation of the format itself?
    - Or is there a need to provide better guidance to item writers (coach writers with exemplars of MC items that do get at higher-order thinking skills)?
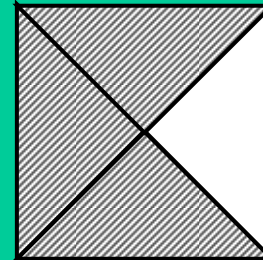
Multiple-choice (Example 1)
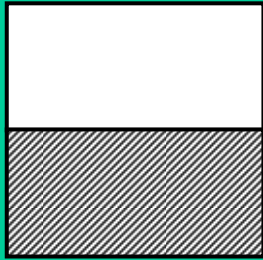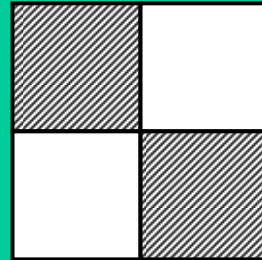
A   B   C   D

Which of the four pictures above shows 3/4?

(Choose one answer.)

1. A

2. B

3. C

4. D

Variation on Multiple-Choice: AICPA

Conceptualizing Item Types

- Continuum from Bennett, Ward, Rock, and LaHart (1990)

  – Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)

Multiple choice		Reordering/ rearrangement		Completion		Presentation
	Selection/ identification		Substitution/ correction		Construction	

Selection/identification

- Similar to MC in that examinees choose a response from presented alternatives

- Number of possible choices is typically large enough to limit possible effects of guessing
  - Some conceptual examples include multiple selection, matching, and cloze elide

Multiple Selection: AICPA

Multiple Selection: AICPA

Highlighting Text
(Carey, 2001; Walker & Crandall, 1999)

"Examinee ability estimates are not group-dependent. Ability estimates obtained from different sets of items will be the same (except for measurement errors), and item parameter estimates obtained in different groups of examinees will be the same (except for sampling errors). In item response theory, item and ability parameters are said to be *invariant*. The property of invariance of item and ability parameters is obtained by incorporating information about the items into the ability-estimation process and by incorporating information about the examinees' abilities into the item-parameter estimation process."

[From: Hambleton, R.K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, NJ: Sage.]

Click on the sentence in the paragraph to the left that *best* describes the practical meaning of the property of invariance.

(Place the pointer on the sentence you wish to highlight, and click to select it.)

? Help

Previous

Next

Multiple Selection

Marco has $7.00 to spend on his lunch. Click on a drink, a sandwich, and a side that Marco could have for lunch so that the price of the three items adds up to $7.00 or less.

Drinks		Sandwiches		Sides	
$1.25	Milk	$3.75	Veggie	$1.50	Chips
$.95	Tea	$4.50	Turkey	$1.00	Cookies
$1.50	Soda	$4.75	Roast Beef	$1.75	Rice
$1.25	Juice	$5.00	Pastrami	$1.50	Green Salad
$.95	Coffee	$4.75	Ham	$1.50	Fruit Salad

Analyzing Situations
(Ackerman, Evans, Park, Tamassia, & Turner, 1999)

© Center for Educational Assessment, 2005

Image courtesy of ETS/C. Mills (2000)

Specifying Relationships
(Fitzgerald, 2001)

For each pair of fractions listed below, decide if the quantities are equal or if they are not equal.

Use the symbols below to replace the question mark in each pair with the correct relationship.

$$\frac{3}{6} = \frac{1}{2} \qquad \frac{1}{3} \; ? \; \frac{4}{6}$$

Drag this symbol to show that the fractions **are** equal

$=$
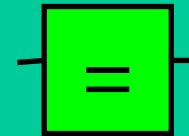
$$\frac{4}{5} \; ? \; \frac{7}{8} \qquad \frac{2}{8} \; ? \; \frac{1}{4}$$

Drag this symbol to show that the fractions **are not** equal
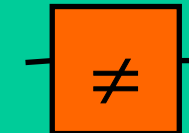
$\neq$

Conceptualizing Item Types
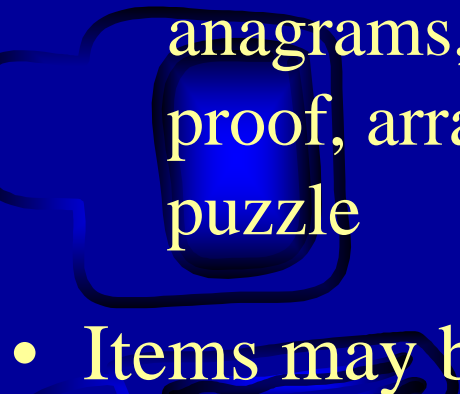
- Continuum from Bennett, Ward, Rock, and LaHart (1990)
  - Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)

Multiple choice

Selection/ identification

Reordering/ rearrangement

Substitution/ correction

Completion

Construction

Presentation
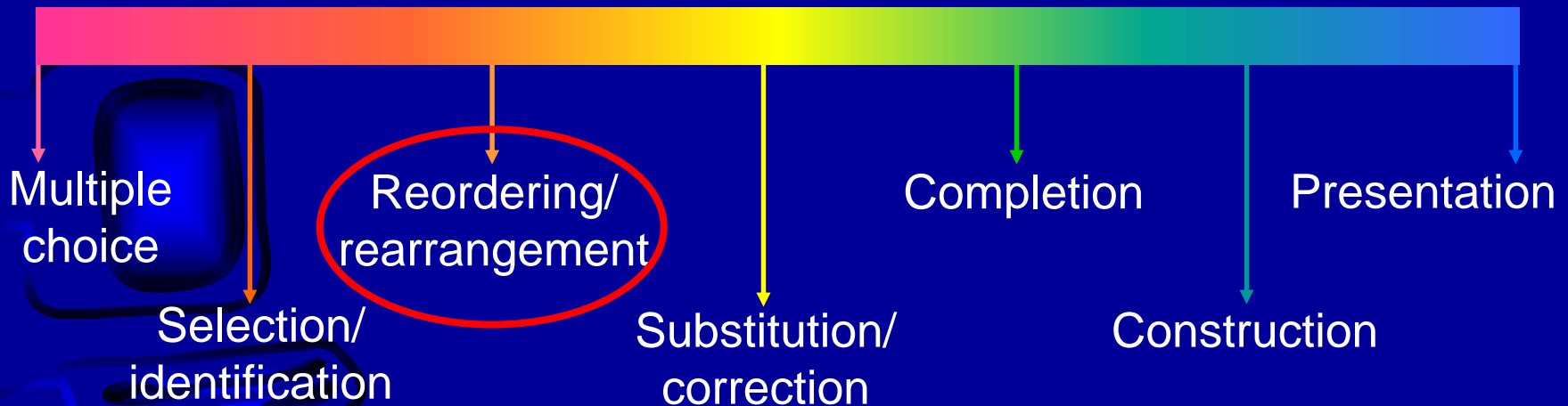
Reordering/rearrangement

- Responses still chosen from stimulus array (presented alternatives/items)

- Task is to place items in correct sequence (or an alternative)

  – Conceptual examples: Ordering sentences, constructing anagrams, arranging mathematical expressions to form a proof, arranging pictures in sequence, putting together a puzzle

- Items may be words, number, pictures, symbols…

Ordering Information
(Fitzgerald, 2001; Walker & Crandall, 1999)

The two-parameter logistic model was fit to five items, and the item parameter estimates for each item are listed in the table below.

Item	a-param	b-param
Item 1	0.45	-0.53
Item 2	1.07	0.09
Item 3	0.84	0.89
Item 4	0.79	0.27
Item 5	1.20	-0.78

Task:

Rank order the five items from the table in terms of their difficulties from hardest to easiest by dragging the item number to the answer box below.

Answer Box

Hard

1. Item 3
2. Item 4
3. Item 2
4. Item 1

Easy 5. Item 5

DONE

Ordering Information
(Microsoft, n.d.)

Beginning

Most insects undergo several distinct changes in form during their development from egg to adult; such a series is called `metamorphosis.~ Each new stage, or `molt,~ occurs when the insect sheds and replaces its outer covering. Some insects, such as grasshoppers, undergo gradual, or `incomplete,~ metamorphosis, with each molt resulting in a more adult-like appearance. Other insects, such as butterflies, moths, beetles, and flies, undergo `complete~ metamorphosis. In complete metamorphosis, the insect passes through four stages of development: (1) egg, (2) larva, (3) pupa, and (4) adult.

The larva is segmented and wormlike. In the case of a butterfly or a moth, the larva takes the form of a caterpillar. The insect eats, often voraciously, during the larva stage.

After a period of feeding and rapid growth, the larva enters the pupa stage. In the pupa stage, the insect encloses itself in a case, or cocoon, and does not

Below are labeled illustrations of the three stages of a moth's metamorphosis. Illustrate the life cycle of the moth by moving the appropriate illustrations to their correct locations <u>outside</u> the circle. The labeled illustration of the first stage, eggs, has already been entered. (Place the pointer on the illustration you wish to move and click. Then place the pointer on the location where you want the illustration to appear and click.)

Larva   Adult   Pupa

Adult

Pupa   Eggs

Test   Quit   Time   Review   Mark   Erase   Calc   Help   Prev   Next

Image courtesy of ETS/C. Mills (2000)

Select-and-Classify
(Microsoft, n.d.)

Some of the number sentences below are adding two numbers together, and some are subtracting one number from the other.

For each number sentence, decide if it is addition or subtraction, click on it, and drag it to the appropriate box.

5+6=11	4-2=2	5-1=4	9+4=13	1+5=6
12-7=5	3+7=10	3-2=1	14-9=5	2+6=8

Addition

5+6=11

Subtraction

12-7=5	4-2=2

Select-and-Place
(Fitzgerald, 2001; Microsoft, 1999)

House numbers on Memorial Drive go west to east. Odd numbered houses are on one side of Memorial Drive, and even-numbered houses are on the other.
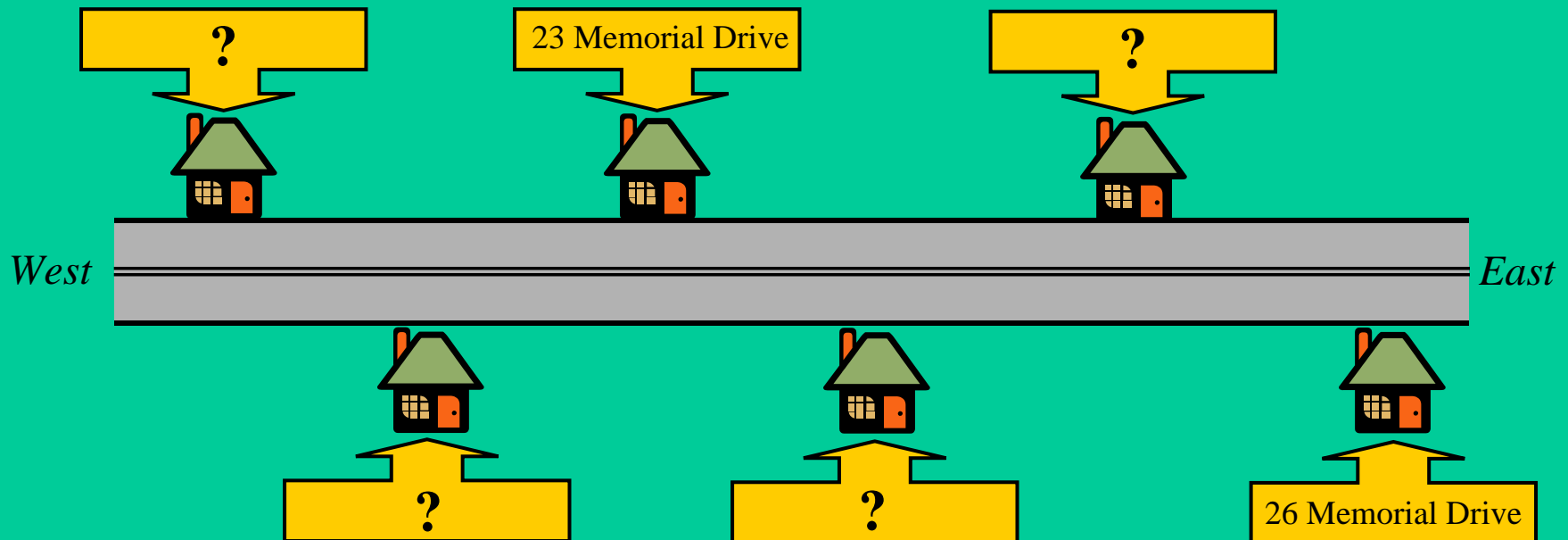
For each address in the table, determine which address corresponds to each house, and drag the address to the appropriate house.

| 22 Memorial Drive | 24 Memorial Drive | 21 Memorial Drive | 25 Memorial Drive |

?

23 Memorial Drive

?

West

East

?

?

26 Memorial Drive

Inserting Text
(Taylor, Jamieson, Eignor, & Kirsch, 1998)

Image courtesy of ETS/C. Mills (2000)

Create-a-Tree (Fitzgerald, 2001; Walker & Crandall, 1999)
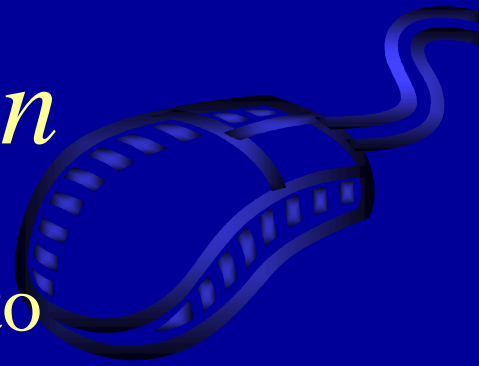
Image from Microsoft (2001)

Conceptualizing Item Types

- Continuum from Bennett, Ward, Rock, and LaHart (1990)

  – Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)

Multiple choice

Selection/ identification

Reordering/ rearrangement

Substitution/ correction

Completion

Construction

Presentation

Substitution/correction

- Examinees must replace (as opposed to reordering/rearranging) what is presented to them with a correct alternative

- Conceptual examples: Correcting spelling, fixing grammatical errors, substituting appropriate words, editing computer programs, formulating mathematical expressions
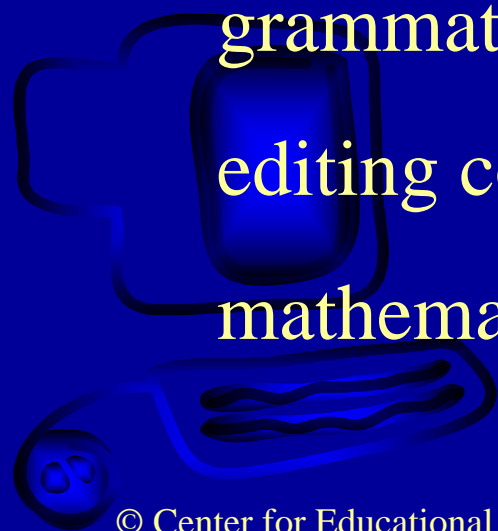
Formulating Mathematical Expressions

The equation for the two-parameter logistic IRT model is:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Using the equation tool below, change the equation given above to correctly represent the mathematical expression for the three-parameter logistic IRT model. (Your answer will appear in the 'My Answer' box after closing Equation Master.)

Equation Master…

Next Item »»➡

My Answer

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Highlighting Text
(Carey, 2001; Walker & Crandall, 1999)

Directions: Click on the <u>one</u> value in the number sentence below that must be changed in order for the number sentence to be correct.

$$3+2+5 = 5+2+9 = 10$$

This is one way to allow students to demonstrate understanding of the commutative property of addition
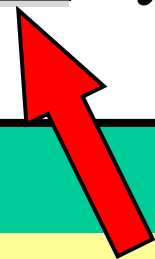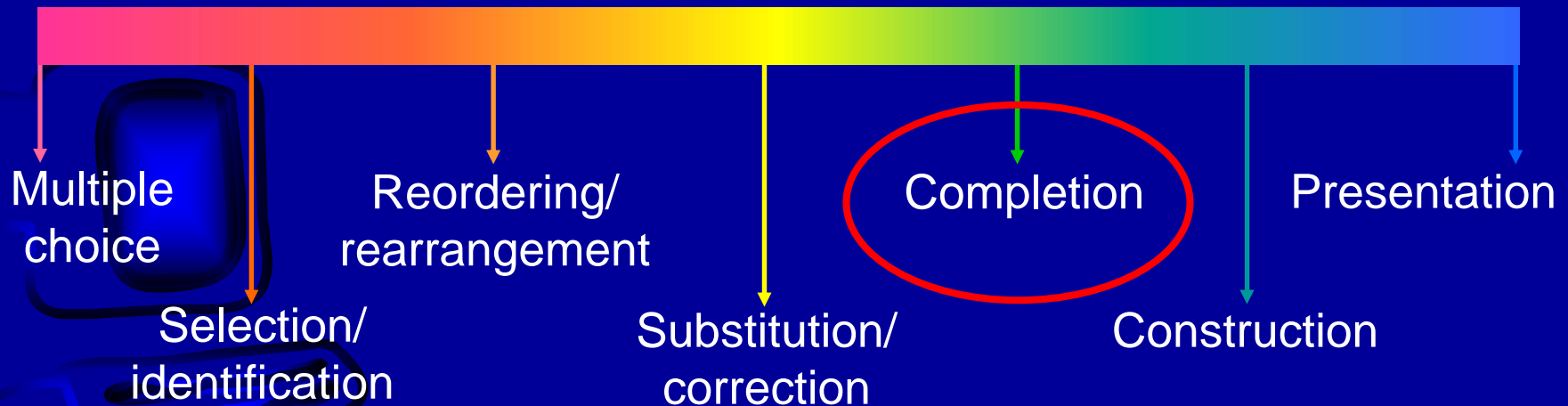
Highlighting Text
(Carey, 2001; Walker & Crandall, 1999)

To find the <u>mean</u> of a list of numbers, you should <u>add</u> all of the numbers and subtract by the number of elements in the list.

Clearly, this should read 'divide'

Conceptualizing Item Types

- Continuum from Bennett, Ward, Rock, and LaHart (1990)
  - Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)

Multiple choice

Selection/ identification

Reordering/ rearrangement

Substitution/ correction

Completion

Construction

Presentation

Completion

- Task is to respond correctly to an incomplete stimulus

- Conceptual examples: Cloze items, sentence completion, mathematical problems with single response, grid-in, short answer

Completion: AICPA

Completion: AICPA

Free-Response Mathematics
(Braswell & Kupin, 1993)

Scores on two items for five examinees are given below. Complete the table by computing the classical item difficulty for each of the items and typing it into the highlighted boxes provided.

(You may use the calculator tool as needed.)

Examinees	Item 1	Item 2
1	0	1
2	1	1
3	1	1
4	0	0
5	1	1
	Difficulty: 0.60	Difficulty: 0.80

CALC   BACK   NEXT
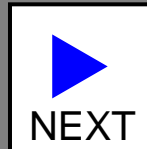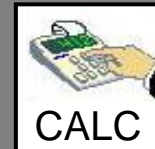
Graphical Modeling
(Bennett, Morley, & Quardt, 2000)
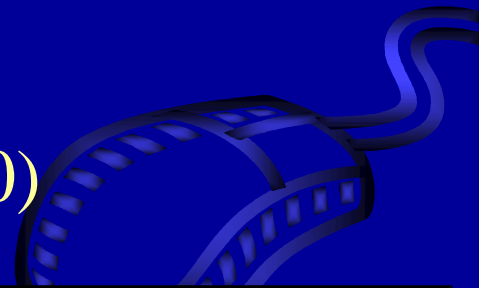
Directions: Use the information in the table below to finish the bar chart. Drag the bar labeled 'Other Color' to show the percent of people in that category.

Eye Color	Percent
Brown Eyes	50%
Blue Eyes	30%
Other Color	20%
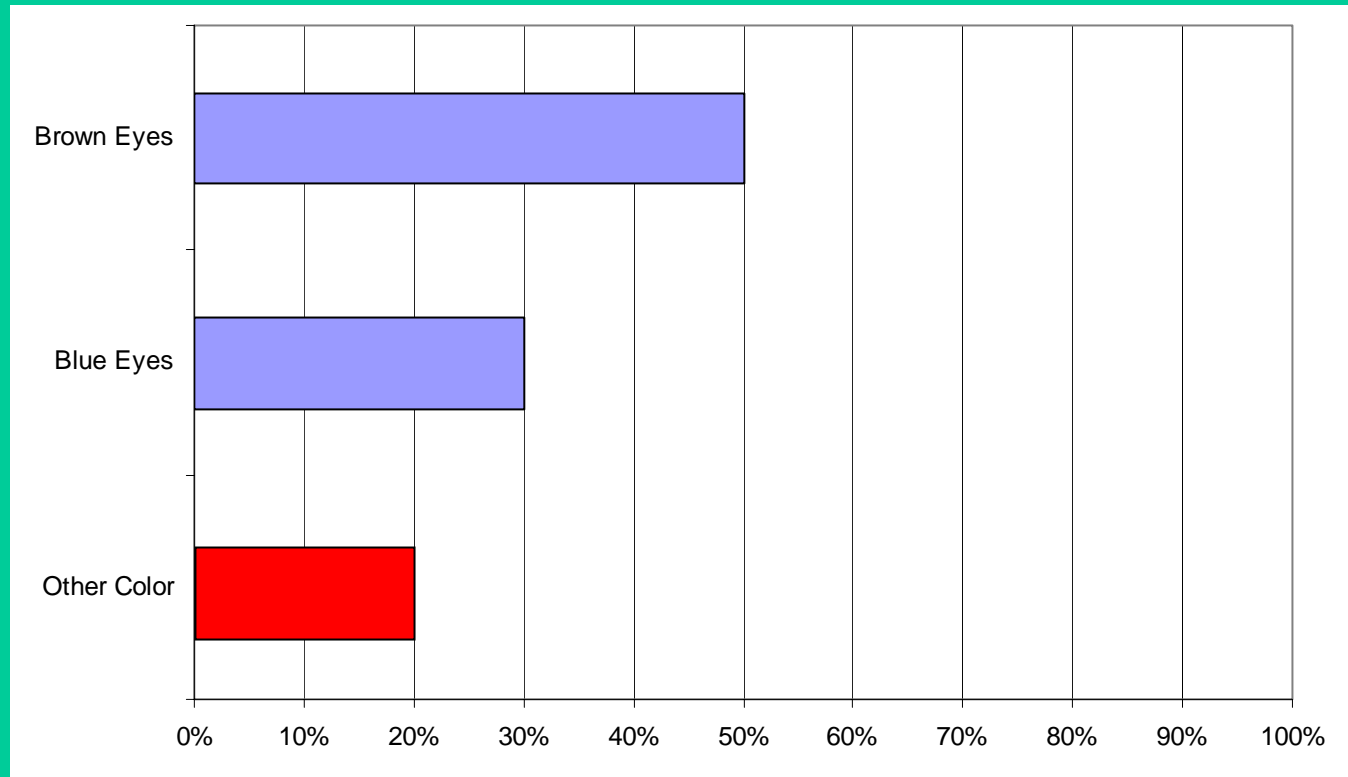
Graphical Modeling
(Bennett, Morley, & Quardt, 2000)
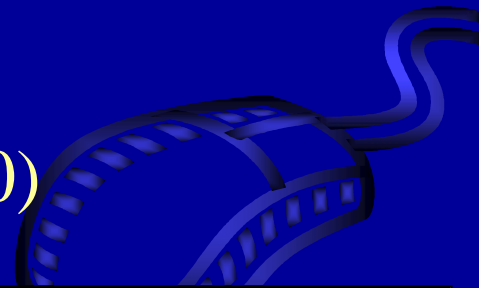
Directions: Shade the area of the figure below to show 50% shaded.
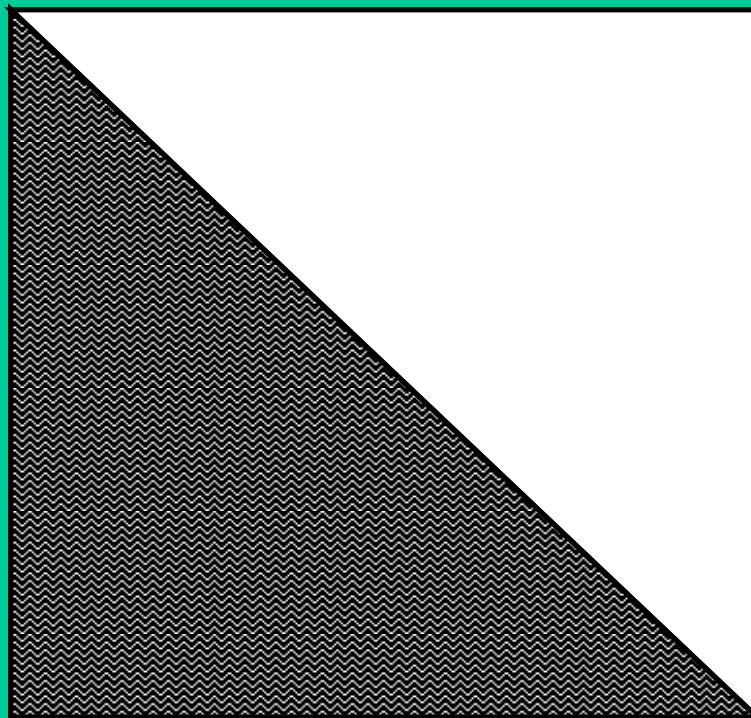
Formulating Hypotheses
(Bennett & Rock, 1995)

Image from ETS (2001)

Numerical Equations (ETS, 2000)

Image courtesy of ETS/C. Mills (2000)

Multiple Numerical Response
(Mills, n.d.)

Miriam is in a play at her middle school, and Miriam's mother wants to buy tickets so that the family can go to the play. Miriam's mother needs to buy tickets for

- Herself
- Miriam's great-grandmother and great-grandfather
- Miriam's two little sisters

Complete the ticket request form below.

How much will it cost for Miriam's family to go to the play?

White Brook Middle School Spring Play

# of tickets		Amount
__1__	adults at $4.00 per ticket……………………..	$4.00
__2__	children at $1.50 per ticket…………………..	$3.00
__2__	senior citizens (55+) at $2.00 per ticket…….	$4.00
	Total Amount	$11.00

Conceptualizing Item Types

- Continuum from Bennett, Ward, Rock, and LaHart (1990)

  – Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)

Multiple choice

Selection/ identification

Reordering/ rearrangement

Substitution/ correction

Completion

Construction

Presentation
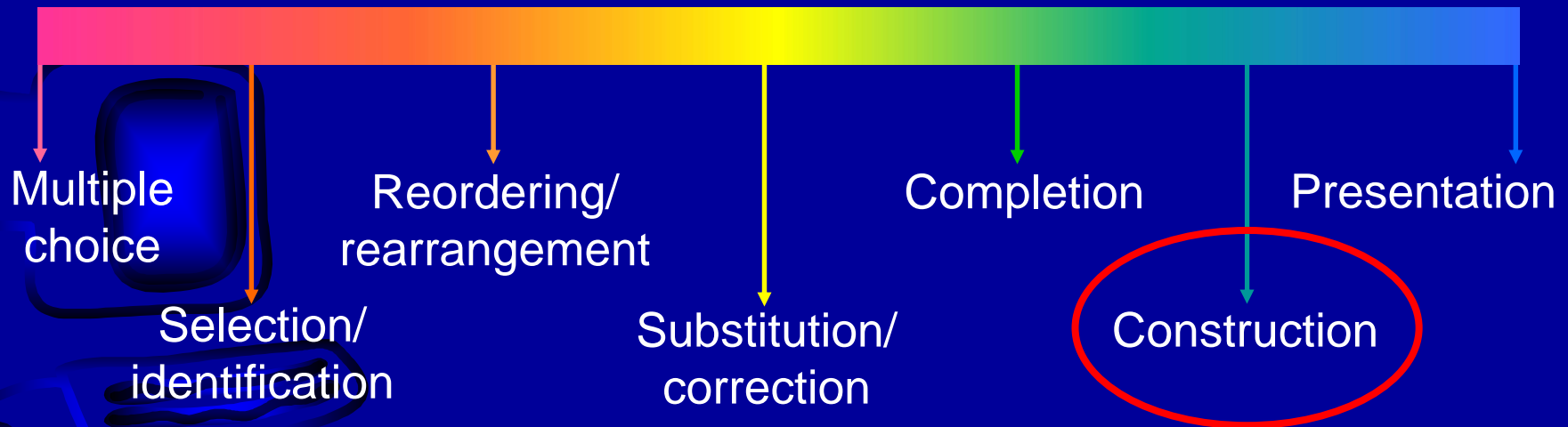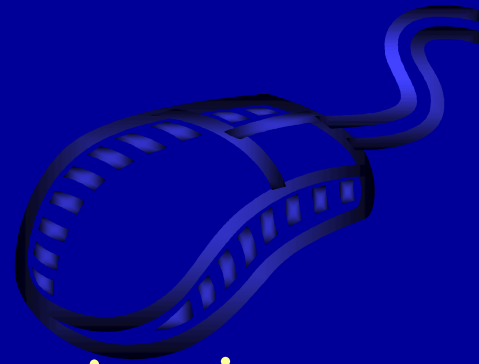
Construction

- In contrast to completion, *construction* items require the entire unit to be constructed

- Conceptual examples: Drawing a complete graph, listing items, essays, mathematical proofs, portfolios, research paper
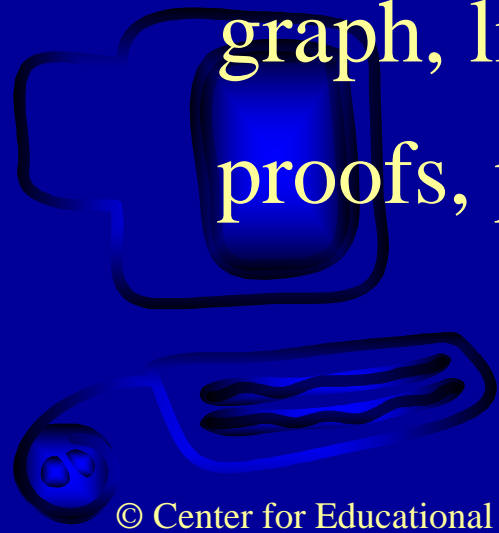
Computer-based essay

In the space provided,
please write an essay addressing the topic given below.

"The earliest known evidence of testing dates back to Chinese employment testing thousands of years ago. Today, we are talking about computer-based testing and the Internet. Briefly (yet completely) describe three of the major developments in measurement practice that have taken place in history."

As educational and psychological tests have found new uses in aspects of everyday life, such tests have evolved in many interesting and exciting ways from their earliest beginnings. Three critical advances that have particularly changed measurement practices include (1) the test development efforts related to the Army Alpha tests during the World War I era, (2) the theoretical innovations associated with the conceptualization of item response theory, and (3), the introduction of the computer as a mechanism for delivering tests and recording examinee responses. In this essay, I will detail how such changes have impacted...

Cut

Paste

Undo

Submit

Graphical Modeling
(Bennett, Morley, & Quardt, 2000)
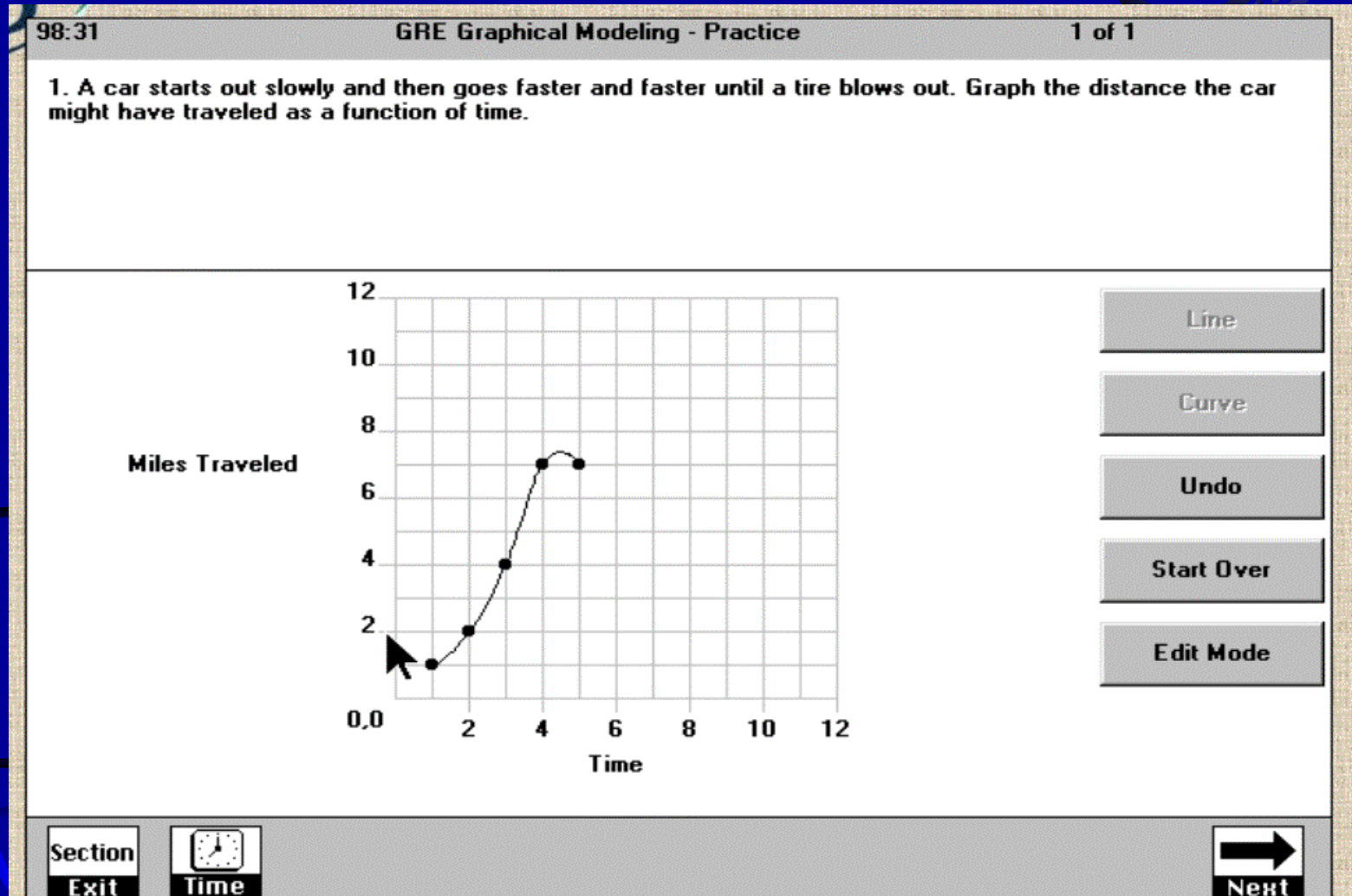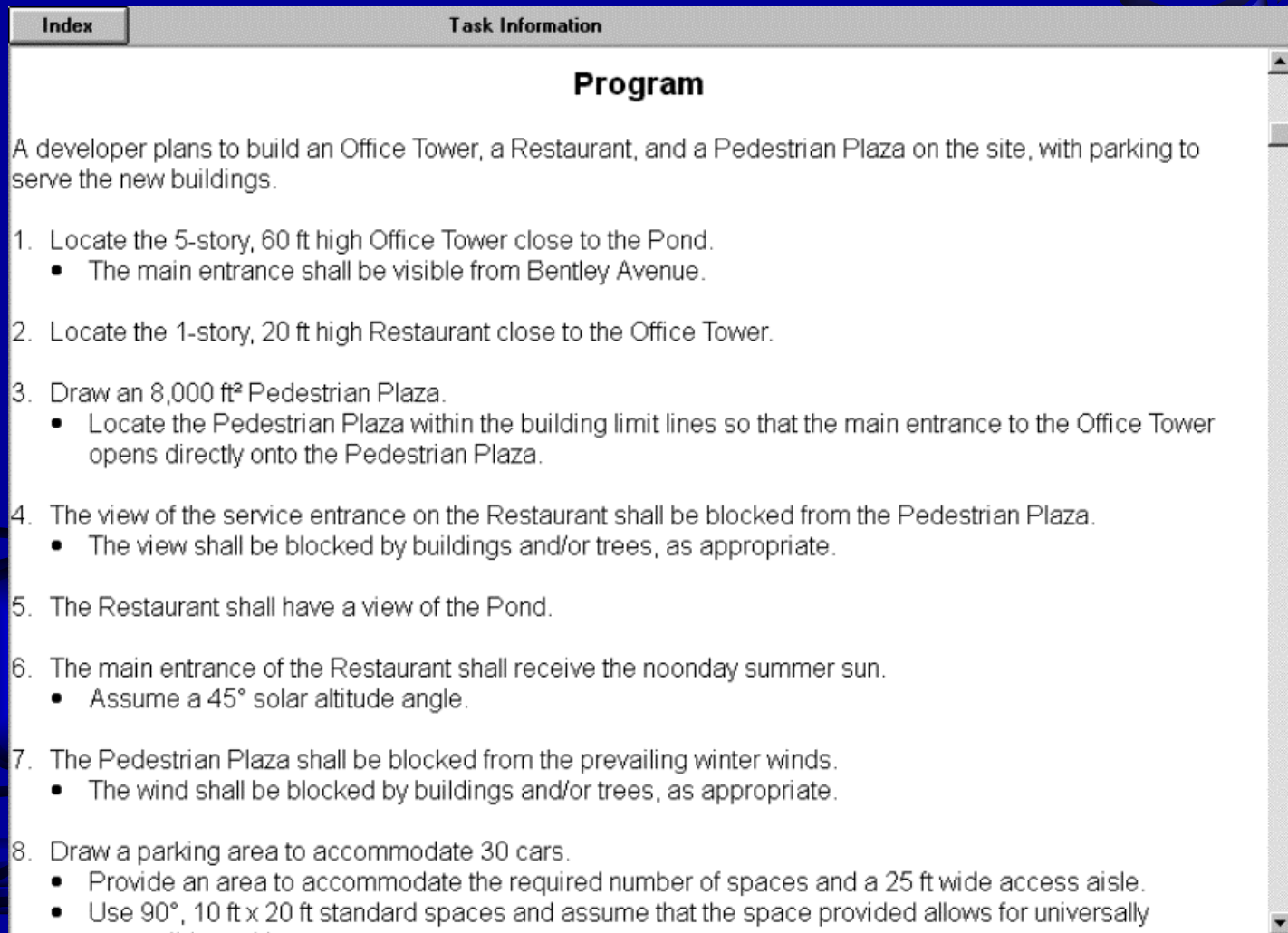
Image from ETS/C. Mills (2000)

Conceptualizing Item Types

- Continuum from Bennett, Ward, Rock, and LaHart (1990)

  – Provides organizing framework for identifying specific cognitive skills and fundamental nature of assessment tasks (in CBT or otherwise)

Multiple choice

Selection/ identification

Reordering/ rearrangement

Substitution/ correction

Completion

Construction
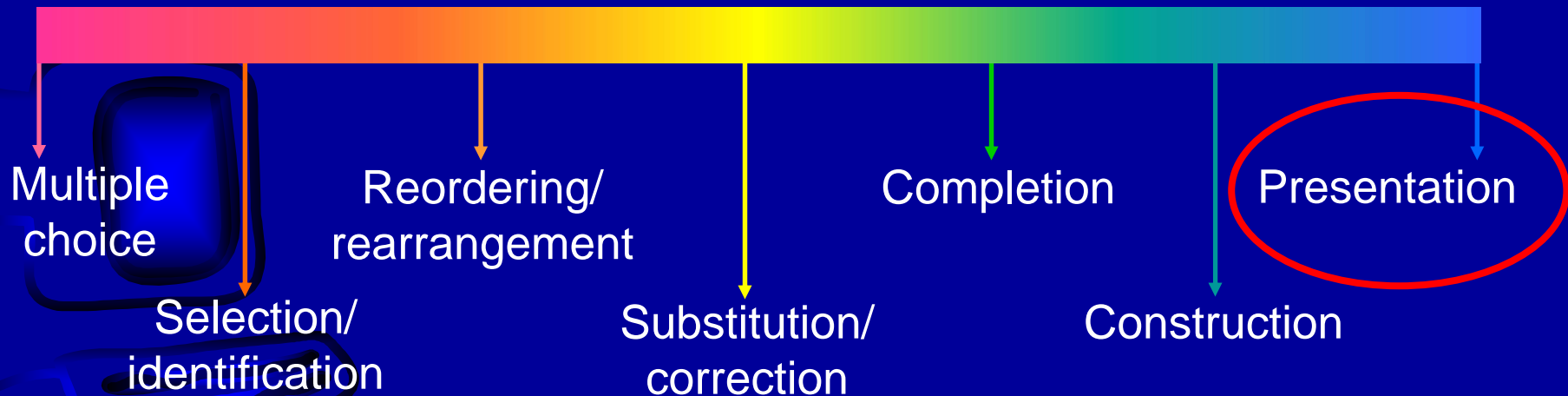
Presentation

Presentation

- Physical presentation or performance delivered under real or simulated conditions in which the object is in some substantial part the manner of performance and not simply its result

- Conceptual examples: performance assessments like laboratory experiments, demonstrations, exhibition, oral presentation, simulations

Site Planning Vignettes (Bejar, 1991)

Program

A developer plans to build an Office Tower, a Restaurant, and a Pedestrian Plaza on the site, with parking to serve the new buildings.

1. Locate the 5-story, 60 ft high Office Tower close to the Pond.
   - The main entrance shall be visible from Bentley Avenue.

2. Locate the 1-story, 20 ft high Restaurant close to the Office Tower.

3. Draw an 8,000 ft² Pedestrian Plaza.
   - Locate the Pedestrian Plaza within the building limit lines so that the main entrance to the Office Tower opens directly onto the Pedestrian Plaza.

4. The view of the service entrance on the Restaurant shall be blocked from the Pedestrian Plaza.
   - The view shall be blocked by buildings and/or trees, as appropriate.

5. The Restaurant shall have a view of the Pond.

6. The main entrance of the Restaurant shall receive the noonday summer sun.
   - Assume a 45° solar altitude angle.

7. The Pedestrian Plaza shall be blocked from the prevailing winter winds.
   - The wind shall be blocked by buildings and/or trees, as appropriate.

8. Draw a parking area to accommodate 30 cars.
   - Provide an area to accommodate the required number of spaces and a 25 ft wide access aisle.
   - Use 90°, 10 ft x 20 ft standard spaces and assume that the space provided allows for universally

Image from NCARB (2000)

Site Planning Vignettes (Bejar, 1991)

Image from NCARB (2000)
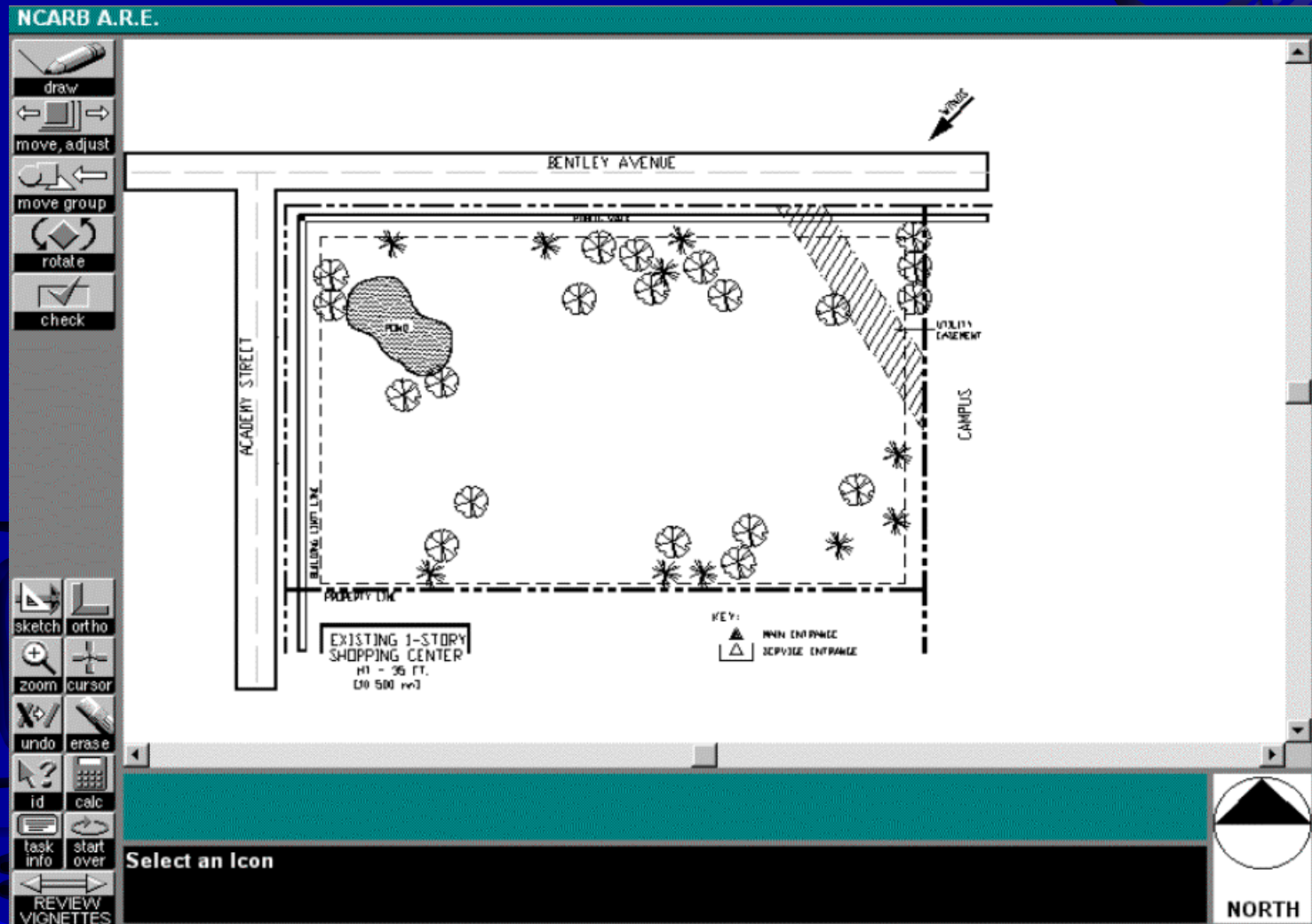
Dynamic Problem Solving Simulation
(Clauser, et al., 1997)

File   Help

| Interval Hx or PE | Write Orders or Review Chart | Obtain Results or See Patient Later
Day 1 @ 16:00 (Wed) | Change Location
Emergency department |

Initial history

Initial history
Reason(s) for Visit:
Chest pain; respiratory distress

History of Present Illness:
The patient, a 65-year-old accountant, was brought to the emergency department by ambulance from the trucking company where he works. Oxygen was administered during transport. About 10 minutes before arrival, he developed excruciating, sharp pain in the right side of his chest and marked respiratory distress. The pain increases with respiration. He is unable to answer questions. A coworker who accompanied the patient to the hospital says that this never happened before, but the patient has had emphysema and asthma for years.

All other history unobtainable.

OK

Image from NBME (2001)

Problem Solving Vignette
(e.g., Microsoft, NCARB)
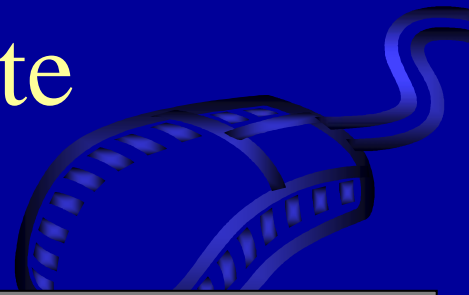
You have been hired as a psychometrician for a major testing company. For your first assignment, you are asked to evaluate the fit of the three-parameter logistic model to some dichotomous test data and report your results to a senior research scientist.

Use the clickable tools below to work through the problem.

When you are finished, click NEXT.

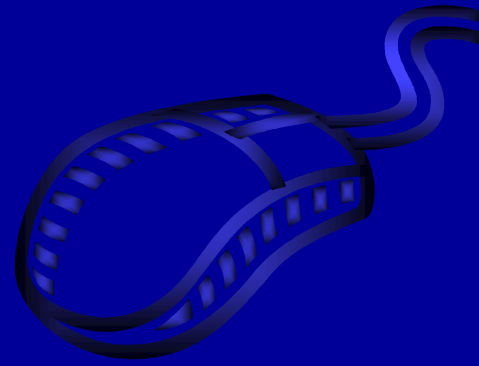|		
Background Information	Look at Data	Choose Analyses
Examine Results	Graphs and Tables	Type Up Findings

Time Remaining: 00:05:25

? Help   ▶ Next

The Need for Tutorials

Technology-Based Assessment Tutorial
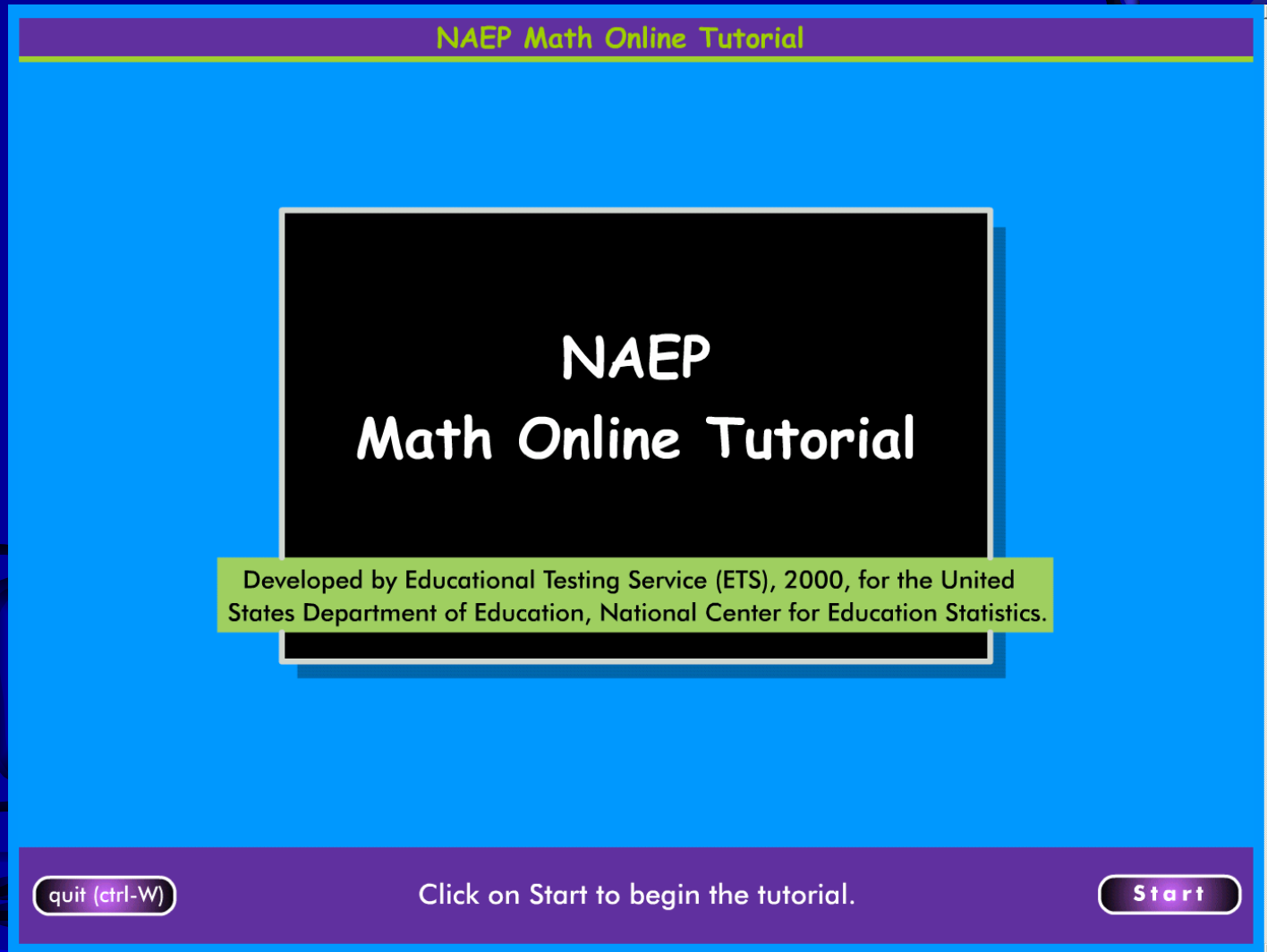
NAEP
Math Online Tutorial

Developed by Educational Testing Service (ETS), 2000, for the United States Department of Education, National Center for Education Statistics.

quit (ctrl-W)   Click on Start to begin the tutorial.   **Start**

Practicing for a CBT

✔ **Scrolling - Quiz**

Directions: Scroll down until you can see "Line 9."

```
Line 4
Line 5
Line 6
Line 7
Line 8
Line 9
```

Yeah, Baby!

quit (ctrl-W)

Click on Go On to continue.

Go On ▶

Using the Keyboard

✔ **Typing Quiz - Part 2**

Directions:

Change the sentence below to read: Max is a **very** happy dog.

Work as quickly as you can. You will have 2 minutes.

Max is a happy dog.

quit (ctrl-W)   When you are finished, click on Go On.   Go On

Intro to Formats

Click on a choice

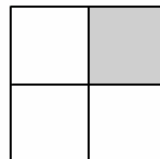When you are answering test questions, most of the time you will click on an oval or words next to the oval.

Washington, D.C.

Sometimes you will be asked to click on a line.
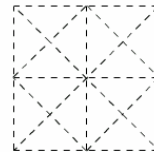
The line you click on will darken. To erase it, click on the line segment again.

Sometimes you will be asked to click on a box.

The box you click on will darken or will be marked with an X. To erase it, click on the box again.

quit (ctrl-W)

Click on Go On to continue.

Go On

Multiple Response Directions

Type a short answer

Directions: Practice typing any word or number into the boxes.
Click on Go On when you are ready to continue.

Answer: 42

Explanation

It's the answer to life, the universe, and everything.

quit (ctrl-W)

Click Go On to continue.

Go On

Using the Calculator Online

✓ Using the Calculator - Quiz

Directions: Use the calculator to solve this problem

12 x 37.5 =

450

+/−	√x	%	÷
7	8	9	×
4	5	6	
1	2	3	
CLR	0	.	=
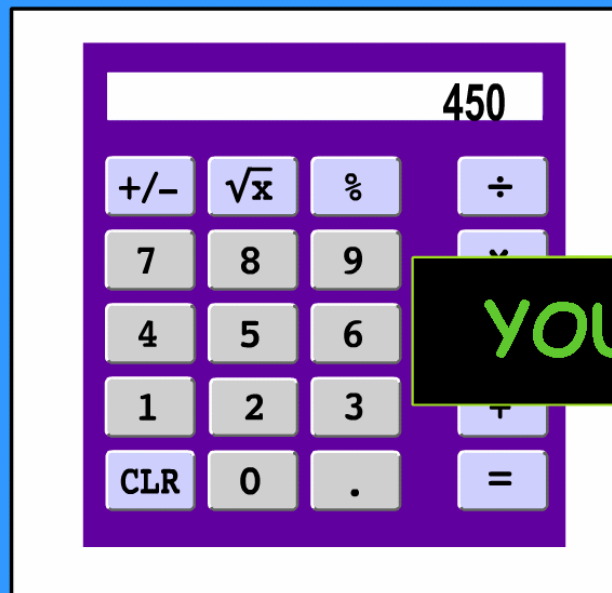
YOU RULE!

quit (ctrl-W)

Click on Go On to continue.

Go On

Other Reference Materials: AICPA

Spell-Check: AICPA

Conclusions and Research Needs

To check out new assessments online:

- www.microsoft.com - (Search for "exam demos")
    - Hot area, active screen, drag and drop, build list and reorder, create a tree, testlet exam format, and simulation question types

- http://www.cpa-exam.org/lrc/exam_tutorial.html
    - Tutorials and sample tests

- http://www.ncarb.org/are/index.html
    - Tutorials and practice software

- http://www.nces.ed.gov/nationsreportcard/studies/tbaproject.asp
    - Overview of computerization with NAEP

- http://134.68.49.185/pegdemo/
    - Submit essay for automated scoring

Testing and Technologies

- Greater use of computerized innovations in CBT has potential to

  – Substantially increase types of skills, abilities, and processes that can be measured

  – Amplify relevance of test tasks to construct

 – Facilitate data collection, management, and scoring

Some Issues
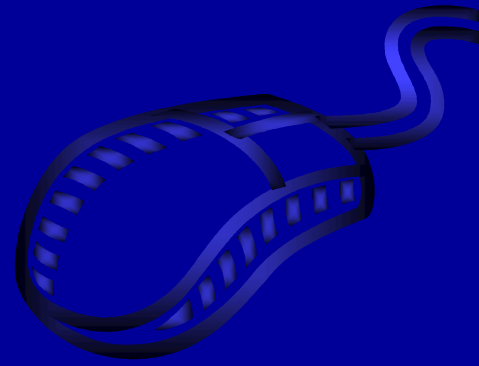
- Obligations of test developers

  1) Construct such tests using appropriate levels of technology given the examinee population and the construct of interest

  2) Explain tasks adequately and unambiguously, both in preparation materials and in pre-test tutorials

- Development costs can be substantial
  - Creating and maintaining item pool

Some Issues

- Scoring innovative items can be complex

- For example, the AICPA

  – Dimensionality

    - Inter-correlations among items

    - Local item dependence within and across tasks

  – Scoring rules

    - Measurement opportunities: Dichotomous or sum up and score polytomously? Or score conjunctively?

Automated Scoring
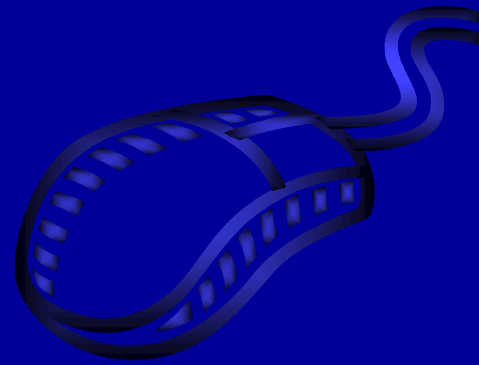
- Why use automated scoring?
  - To implement many of the more novel item types, scoring methods are necessary that can handle responses unique to each examinee in structure and/or nature
  - <u>Not</u> rigid or mechanical application of limited, explicit criteria to examinee responses

- However, any scoring rules should be carefully reviewed…

Example: ETS *Math On-Line*

- *Math On-Line* is CBT tryout of NAEP
  - Sample problem (Gr. 4): Cut a paper tube, what shape is the flattened piece of paper?
    - Respond with text AND/OR draw shape
  - **W I D E** variation in answers observed
    - Spelling, Mismatched answers, odd responses
    - Answers depended on perspective
- Key consideration: level of constraint in constructed-response

Research Needs
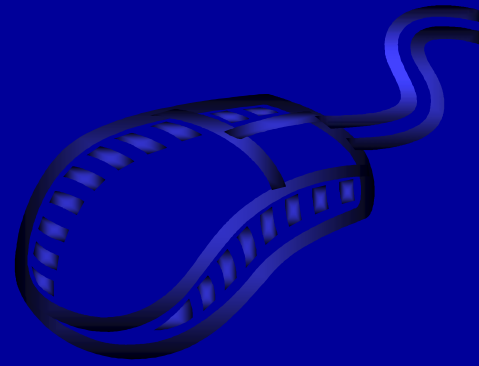
- Pick an area of validity
  - This kind of information on item types in literature is sparse in general; it's possible that in some cases such evidence may be proprietary

- Task specificity (Linn & Burton, 1994)

- Relative complexity/simplicity of interface

- Tutorials and task intricacy

- Task presentation variables

- Human factors (including disabilities)
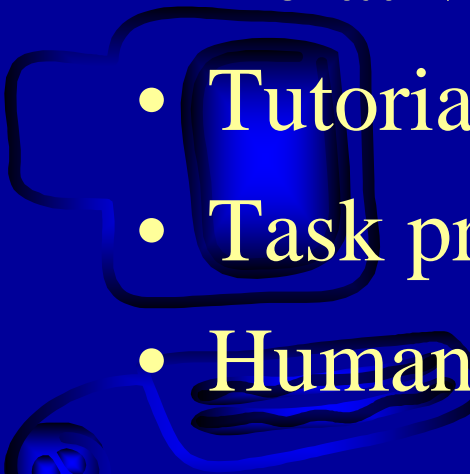
Research Needs
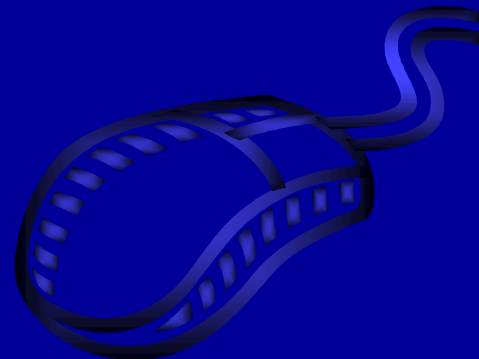
- Potential losses in score validity
  - Selection of scoring algorithm, consistency and generalizability of results
  - "Gaming" of automated scoring systems
    - More studies like Powers, Burstein, Chodorow, Fowles, & Kukich (2001)

- Dimensionality issues
  - Many emerging constructs are multidimensional
  - Multidimensional IRT
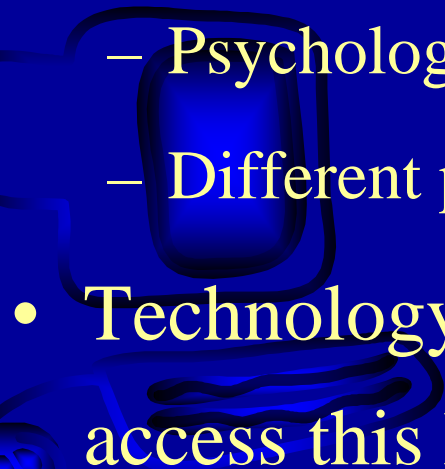   - Dichotomous and Polytomous

Conclusions

- Increased use of computers in testing mirrors escalation of technology in everyday life

- Examinees support opportunity to demonstrate knowledge and skills in ways relevant to and/or analogous to how they use those abilities

Some Cautions

- (This is where the skeptics can come back)

- With research, technology can be a tool for better measurement *in some cases and for some constructs*

- Ultimately, CBT and computer innovations should be integrated with (demonstrable) measurement benefits as the goal

Information has value!

- Information gathered via a test can function differently and provide different benefits (Pearson & Garavaglia, 1997)

  – Decision confidence and measurement precision

  – Novelty (elaboration of new construct or construct dimension)

  – Psychological information (rather than purely statistical)

  – Different perspective on construct

- Technology used with purpose can permit test users to access this kind of information

Potential Constructs

- While recall of facts remains important, there are many possible additional constructs and skills valued depending on context….

Research	Critical Reasoning	Motivation	Teamwork
Leadership	Communication	Analysis	Languages
Information Technology	Judgment	Learning Ability	Organization

Expert Systems Analysis

- Expert systems are programs designed to emulate scoring behaviors of humans

  – *PROUST*; also *MICROPROUST*

  (Braun, Bennett, Frye, & Soloway, 1990)

  – *GIDE*

  (Martinez & Bennett, 1992; Bennett & Sebrechts, 1996)

- These methods generally less able to provide diagnostic information

Mental modeling
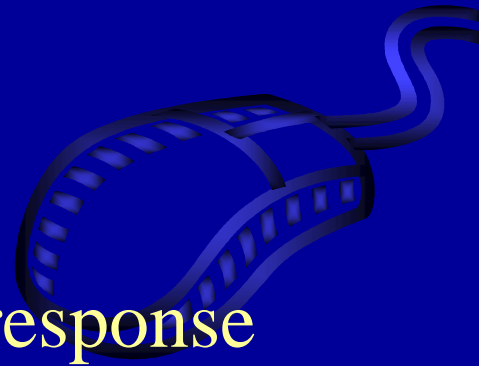
- Elements of examinee's constructed response compared to all possible variations via features extraction analysis programs

  - Mimics scoring of committees; requires consideration of how experienced experts evaluate solutions (Williamson, et al, 1999)

  - Architectural Registration Examination; National Board of Medical Examiners' Step 3 CCS

Multiple-choice (Example 2)

Which of the four pictures below shows 3/4?

(Click on the letter of one answer.)
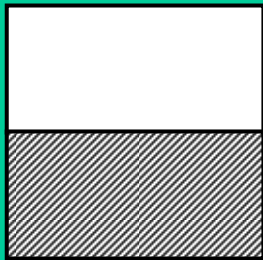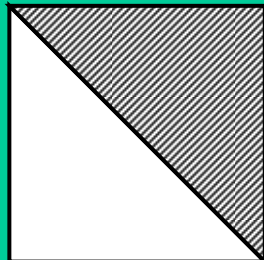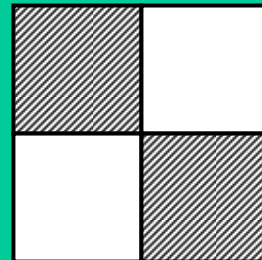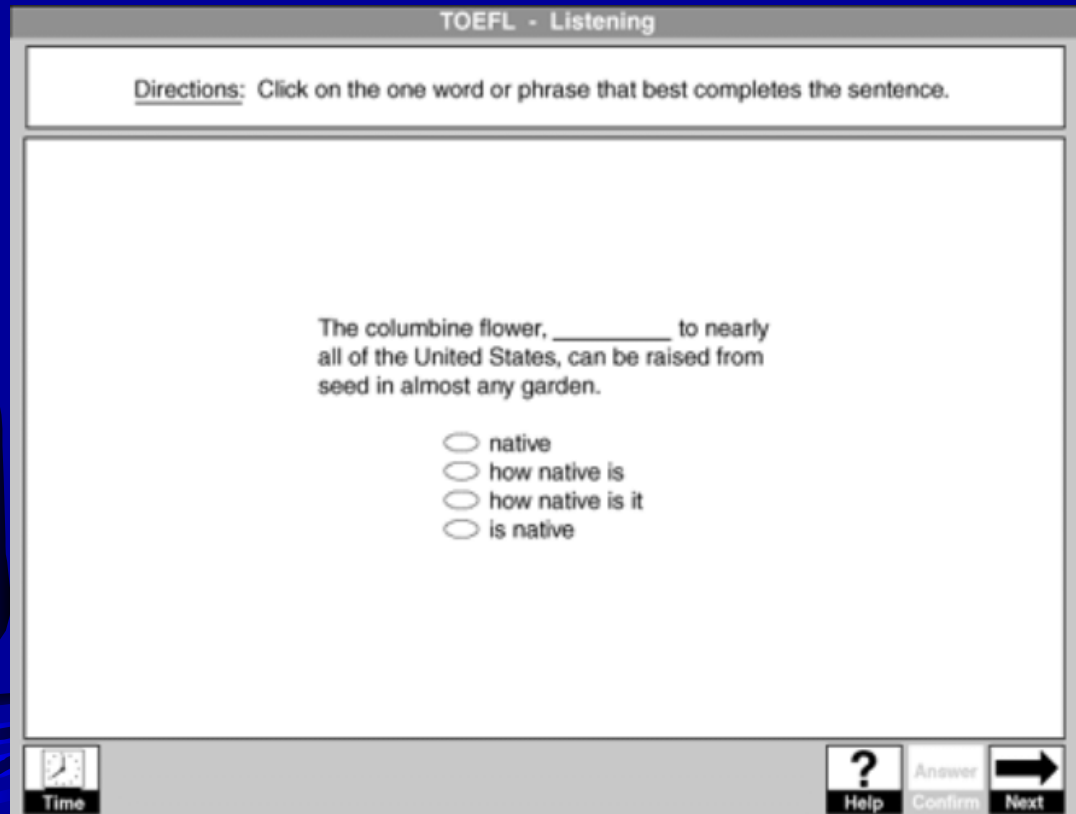
A   B   C   D
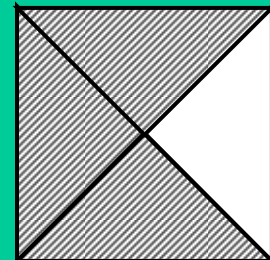
Multiple-choice

- Choose an answer from a small set of response options

TOEFL - Listening

Directions: Click on the one word or phrase that best completes the sentence.

The columbine flower, _____ to nearly all of the United States, can be raised from seed in almost any garden.
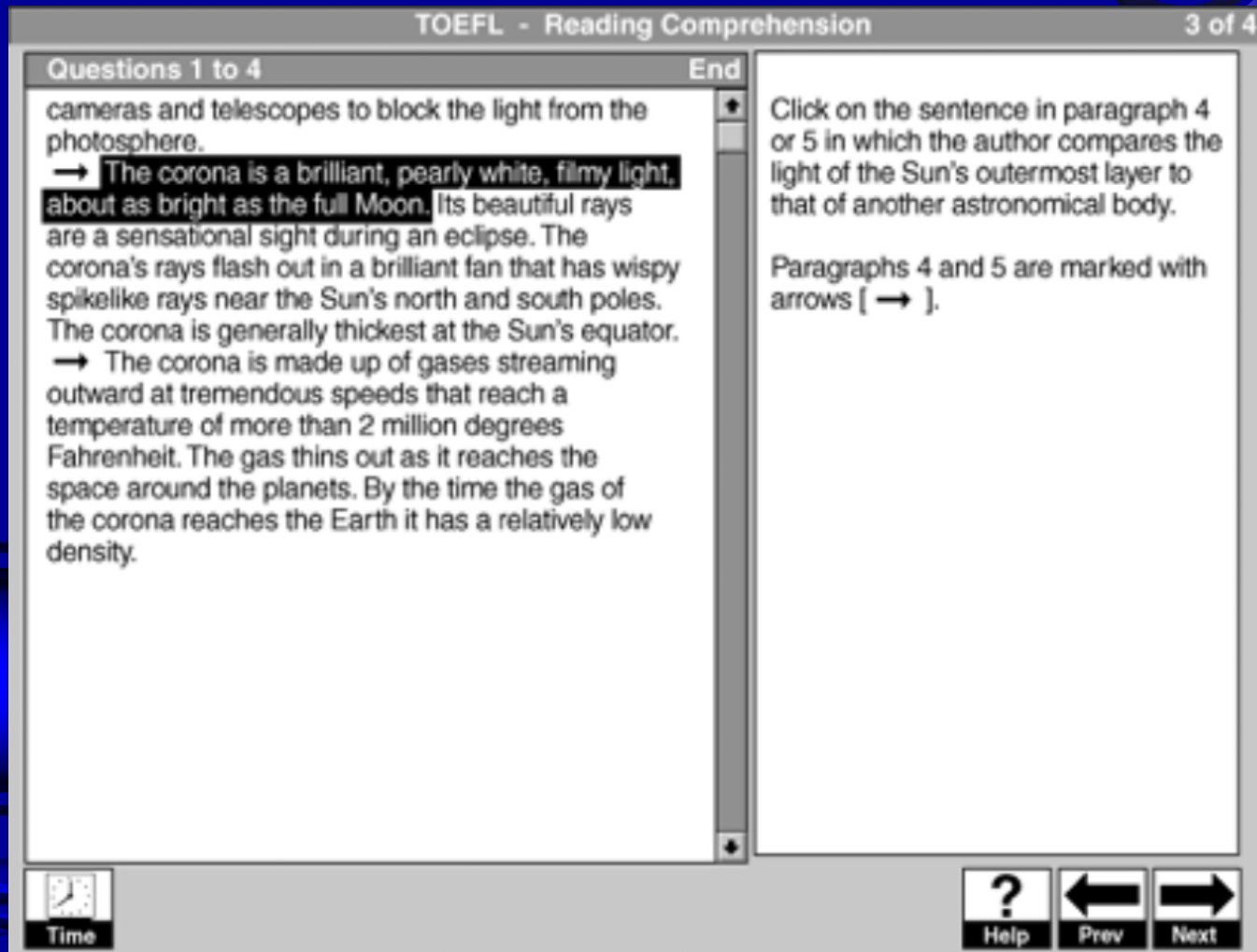
- native
- how native is
- how native is it
- is native

Time | Help | Answer Confirm | Next

Image from ETS (2002)

Highlighting Text
(Carey, 2001; Walker & Crandall, 1999)

Questions 1 to 4   End

cameras and telescopes to block the light from the photosphere.
→ The corona is a brilliant, pearly white, filmy light, about as bright as the full Moon. Its beautiful rays are a sensational sight during an eclipse. The corona's rays flash out in a brilliant fan that has wispy spikelike rays near the Sun's north and south poles. The corona is generally thickest at the Sun's equator.
→ The corona is made up of gases streaming outward at tremendous speeds that reach a temperature of more than 2 million degrees Fahrenheit. The gas thins out as it reaches the space around the planets. By the time the gas of the corona reaches the Earth it has a relatively low density.

Click on the sentence in paragraph 4 or 5 in which the author compares the light of the Sun's outermost layer to that of another astronomical body.

Paragraphs 4 and 5 are marked with arrows [→].

Time

? Help   ← Prev   → Next

© Center for Educational Assessment, 2005

Image from ETS (2002)

Highlighting Text
(Carey, 2001; Walker & Crandall, 1999)

Image courtesy of ETS/C. Mills (2000)

Multiple True-False

Use the mouse to make a ✔ next to each of the shapes below that are examples of <u>polygons</u>. Be sure and mark all of the <u>polygons</u>.
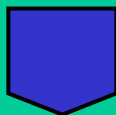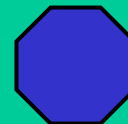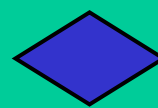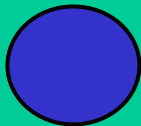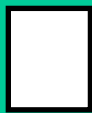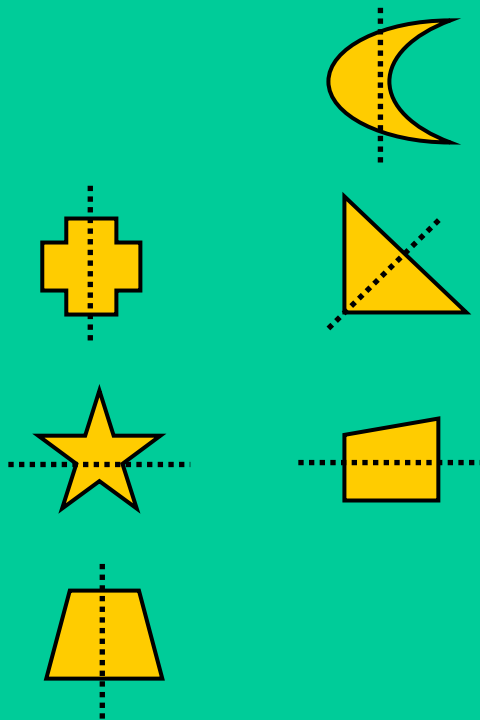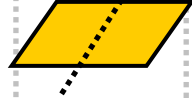
Click here for the next item

Create-a-Tree

(Fitzgerald, 2001; Walker & Crandall, 1999)

Below are several shapes that may or may not be symmetric along the dotted line.

Decide if the shape is symmetric along the dotted line, click on the item, and drag it to the correct box.

Create-a-Tree
(Fitzgerald, 2001; Walker & Crandall, 1999)

Image from Microsoft (2001)

Create-a-Tree (Fitzgerald, 2001; Walker & Crandall, 1999)

Image from Microsoft (2001)

Select-and-Place
(Fitzgerald, 2001; Microsoft, 1999)

Image from Microsoft (2001)
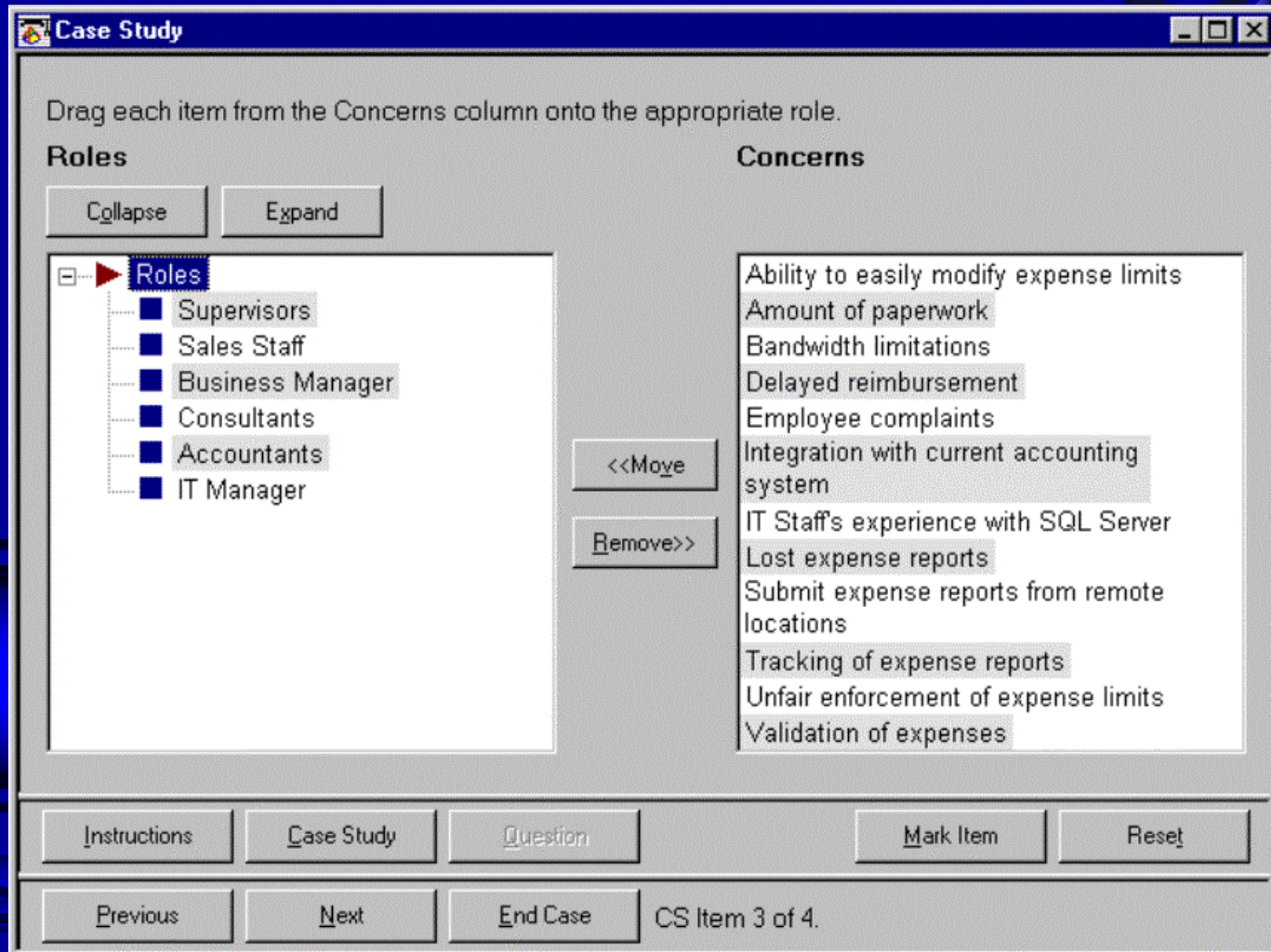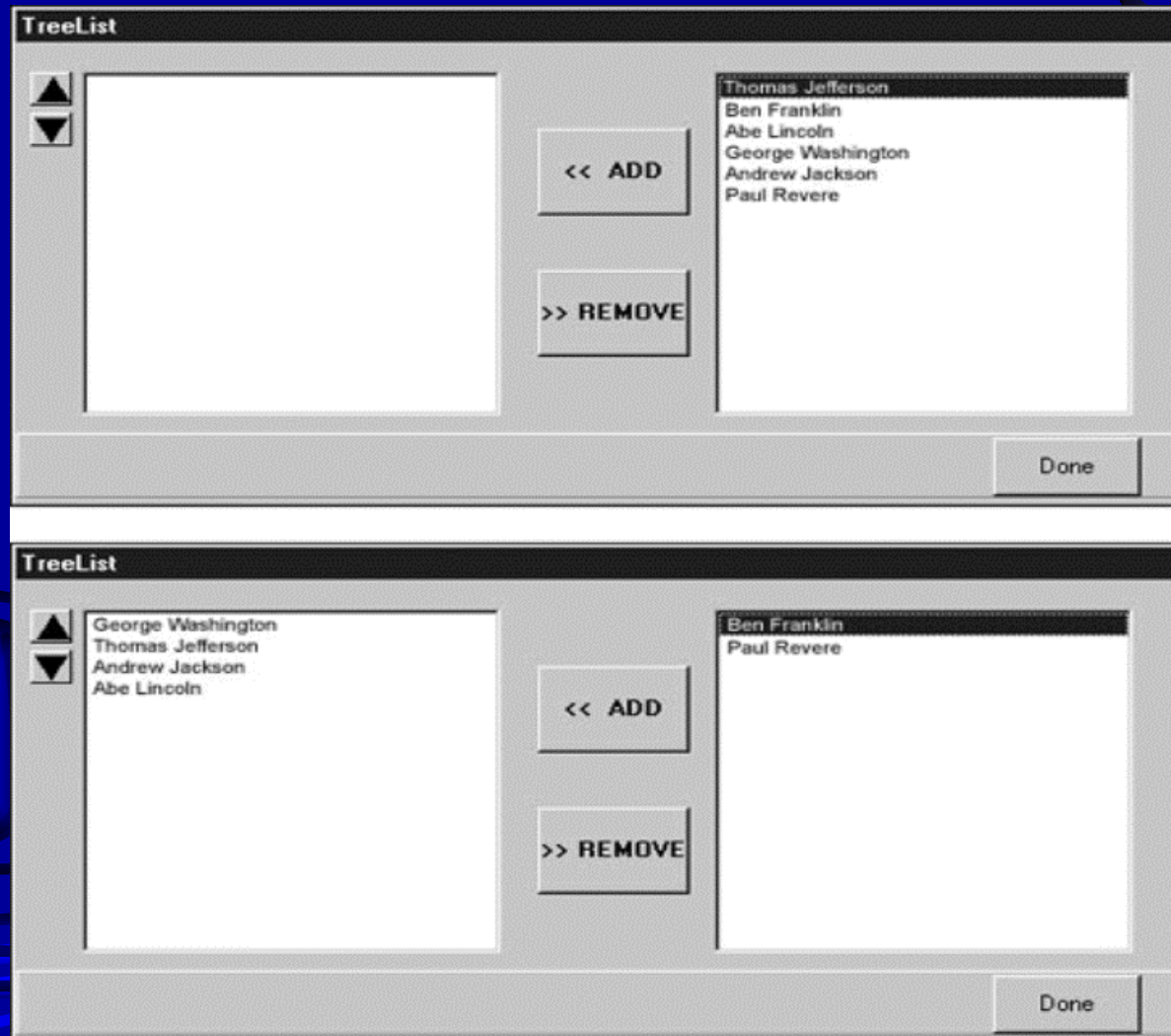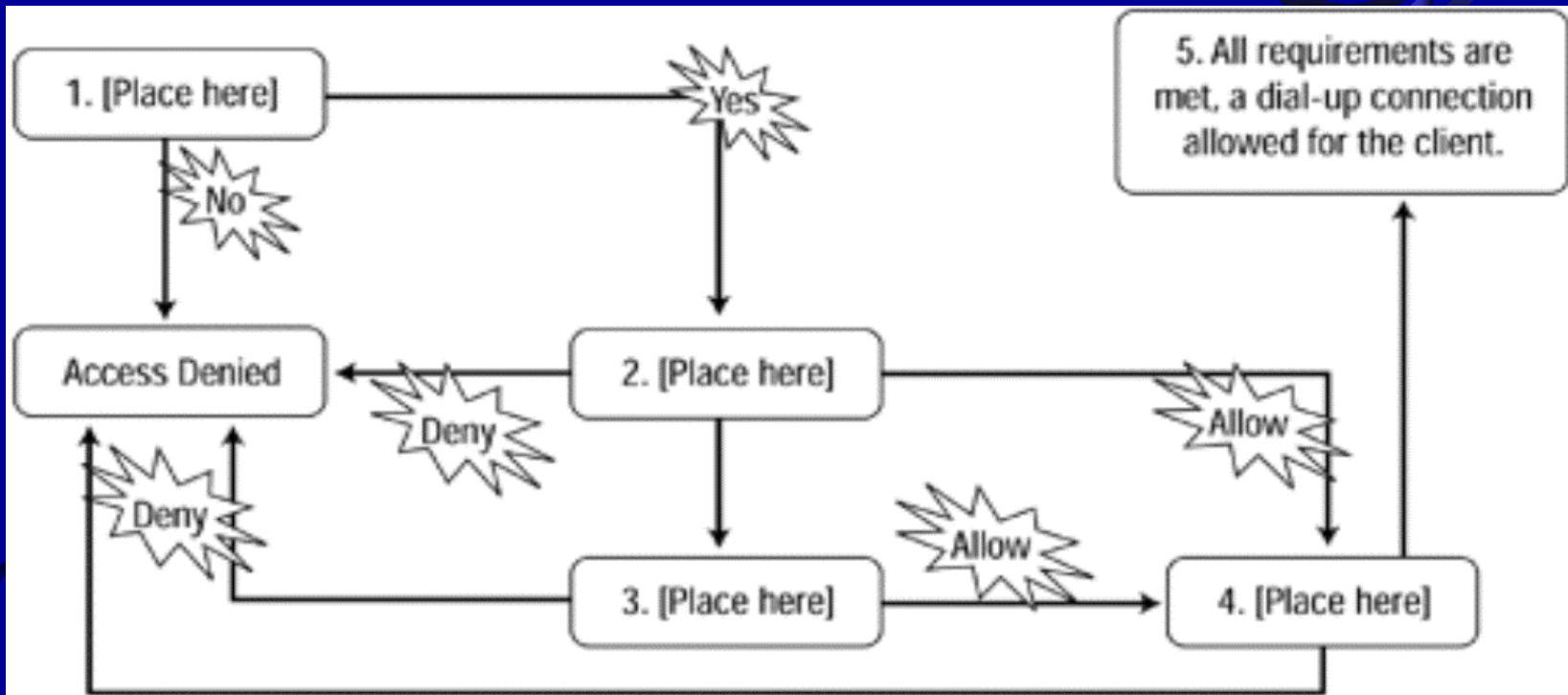
Build, List, and Reorder (Microsoft, n.d.)

Image from Microsoft (2001)
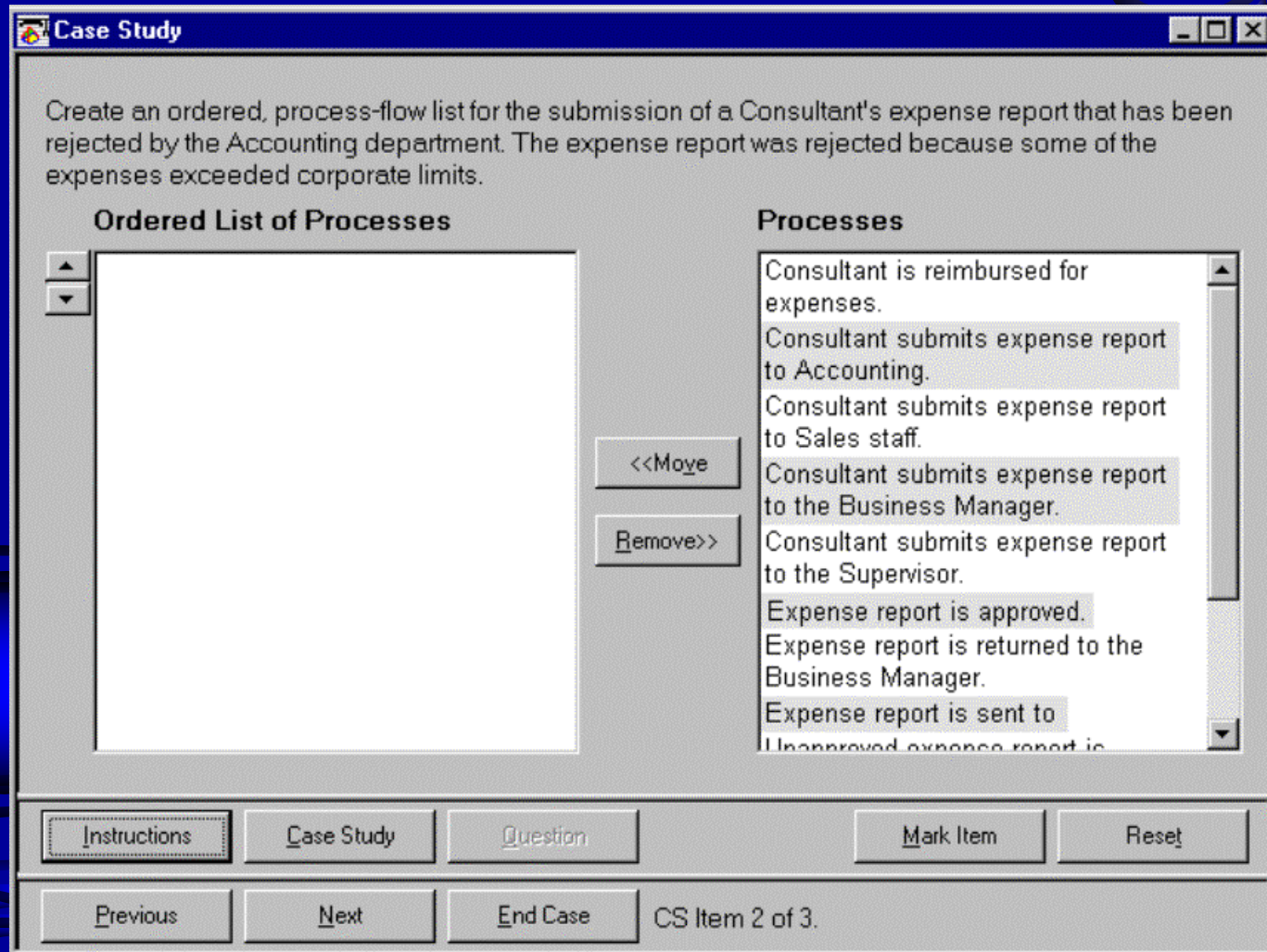
Drag-and-Connect (Fitzgerald, 2001)

You want to visit every planet starting at the planet closest to the sun and ending at the planet farthest from the sun. Indicate your route by connecting the planets in order of your visit. (Use all planets.)

Saturn

Mercury

Jupiter   Neptune   Uranus

>> Route >>

Route

Route

Venus

Mars   Pluto

<< Route <<

Earth

Image from Microsoft (2000)

Sorting Information
(Bennett & Sebrechts, 1997)

Time Remaining: 0:22:07 Problem Sorting

Group A
Tickets to a certain track meet were sold for $4.50 each. If the same number of tickets had been sold for $3.50 each, the revenues from the ticket sales would have been $300 less. How many tickets were sold for the meet?
(a – b)x = c, where a = 4.50, b = 3.50, and c = 300

|2| | | |7| | | | | | | | | | |

Group B
A quarry stocks 2 types of gravel. Type 1 is 100% slate chips by volume and Type 2 is 80% slate chips by volume. If 1,000 cubic feet of Type 2 is mixed with enough Type 1 to make a new mixture that is 40% slate chips by volume, how many cubic feet of Type 1 must be added to make the new mixture?
ab + x = c (x – b), where a = 0.8, b = 1,000, and c = 0.4

| | | | | | | | | | | | | | | |

Card 7
Members of a certain club are selling tickets for a charity event. Tickets cost $3.50, and the members have pledged that out of this only $1.00 will go toward expenses, the rest will go to charity. If they want to give $15,000 to the charity, how many tickets must they sell?

Group C
A rectangular yard is 130 feet long by 110 feet wide. Of this, 2/3 is covered with grass. A walkway covers 1/10 of the area that is not covered by grass and the remaining area contains a patio. What is the area of the patio?
ab(1 – c)² – d(1 – x), where a = 130, b = 110, c = 2/3, and d = 1/10

| |3|4| | | | | | | | | | | | | |

Group D
Tom and Tina are 4 miles apart. Traveling 4 miles takes Tom 14 minutes by bicycle and Tina 50 minutes by foot. If they begin traveling toward each other at these rates, how many minutes would elapse before they meet?
x/a + x/b = 1, where a = 14, b = 50

| | |5|6| | | | | | | | | | | | |

Test
Quit

Click one of the GROUPs to place a CARD. Click on any GROUP to move a CARD. Click on NEXT or PREVIOUS to locate a CARD.

Prev Next

Image courtesy of ETS/C. Mills (2000)

Create-a-Tree (Fitzgerald, 2001; Walker & Crandall, 1999)

Image from Zenisky & Sireci (2002)

Highlighting Text
## (Carey, 2001; Walker & Crandall, 1999)

TOEFL - Listening

Directions: Click on the one underlined word or phrase that must be changed for the sentence to be correct.

One of the <u>most</u> difficult problems <u>in understanding</u>

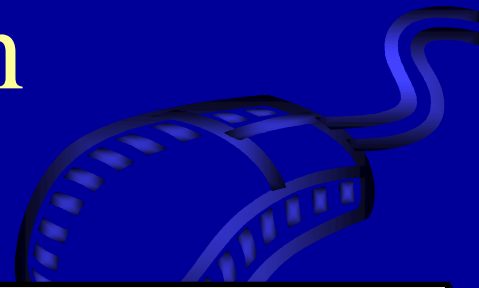sleep is <u>determining</u> what the functions of sleep <u>is</u>.

Time

? Help | Answer Confirm | Next ➡

Image from ETS (2002)

Select-and-Classify (Microsoft, 19XX)

Image courtesy of ETS/C. Mills (2000)
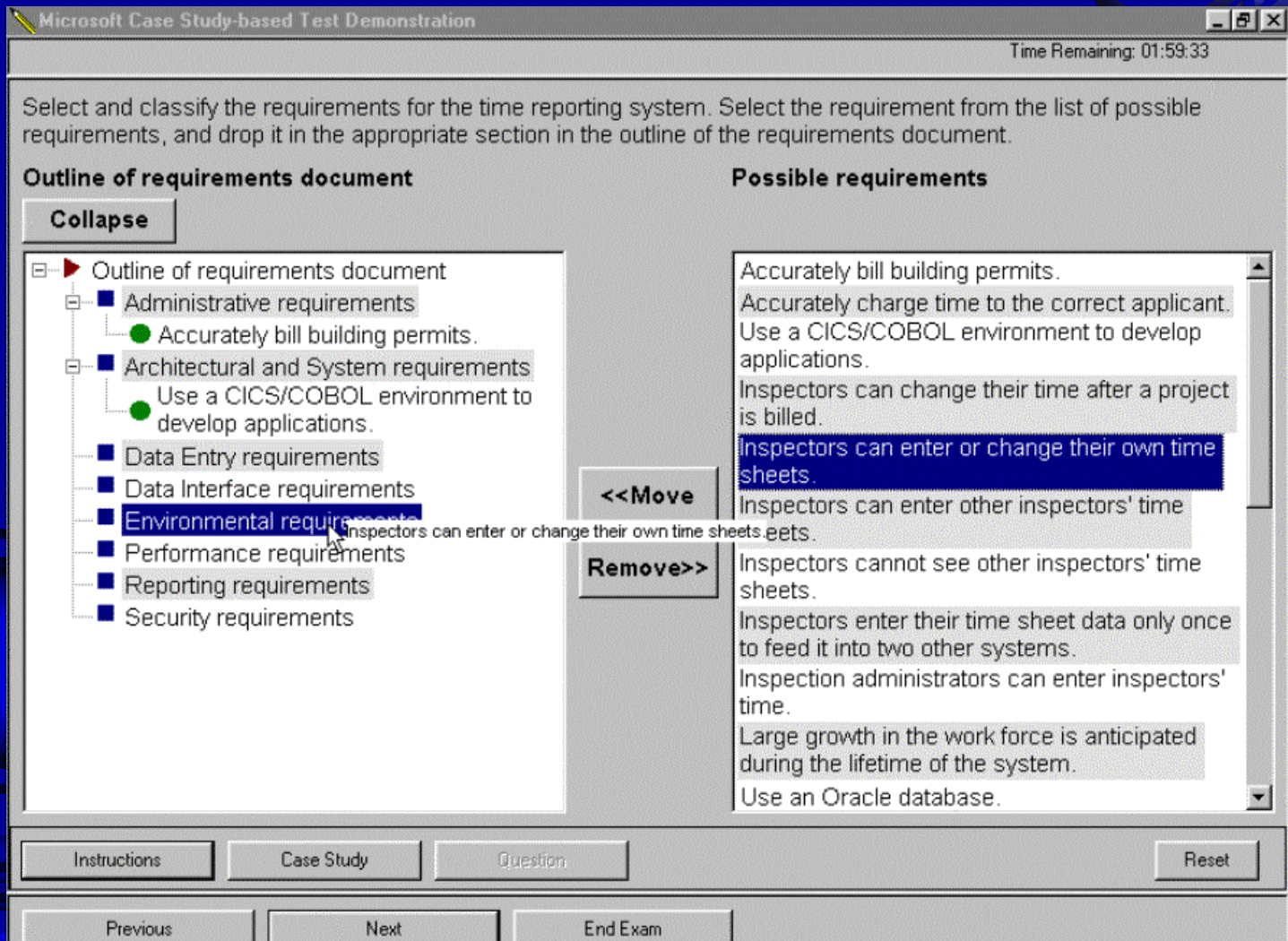
Ordering Information
(Microsoft, n.d.)

Below are listed six cities in Massachusetts and their populations from the last state census.

Re-order the list from the city with the most people to the city with the fewest.

	Most	Springfield: 152,082
Fall River: 91,938		
Pittsfield: 45,793		
Springfield: 152,082		
Lowell: 105,167		
Dedham: 23,464		
Boston: 589,141	Fewest	Dedham: 23,464

Specifying Relationships
(Fitzgerald, 2001)

Image courtesy of ETS/C. Mills (2000)
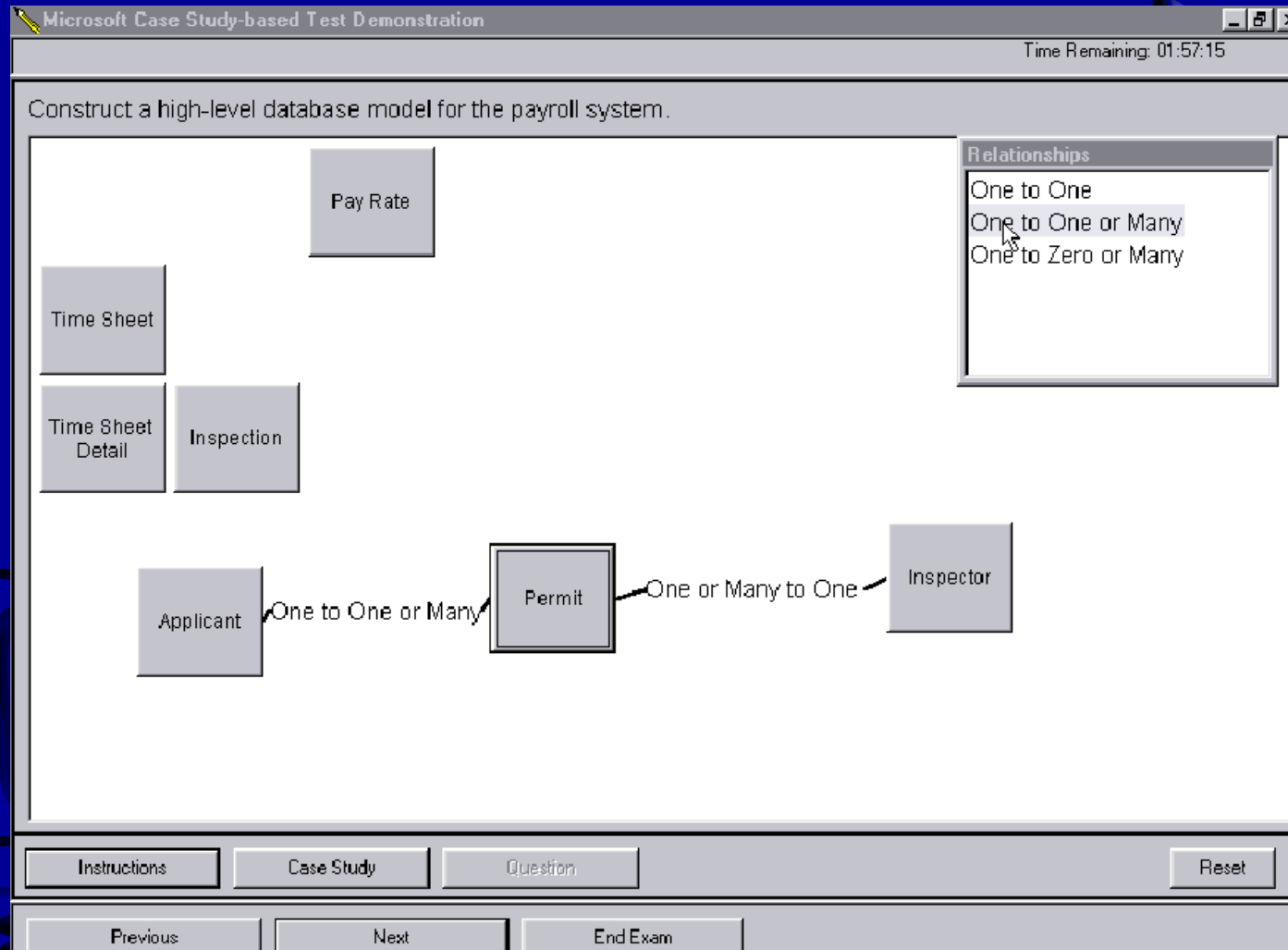
Simple Concept Maps (Novak, 1990)
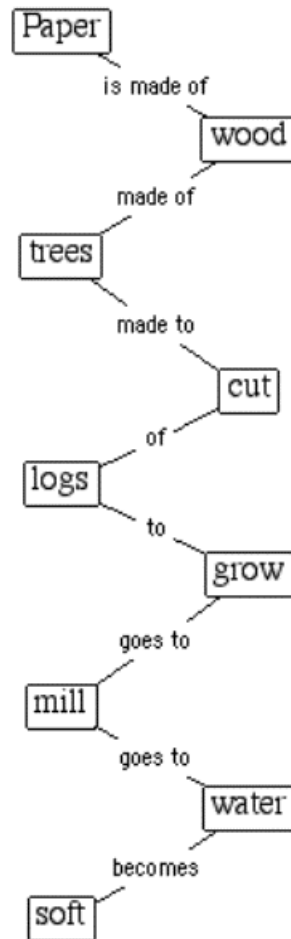
1

2
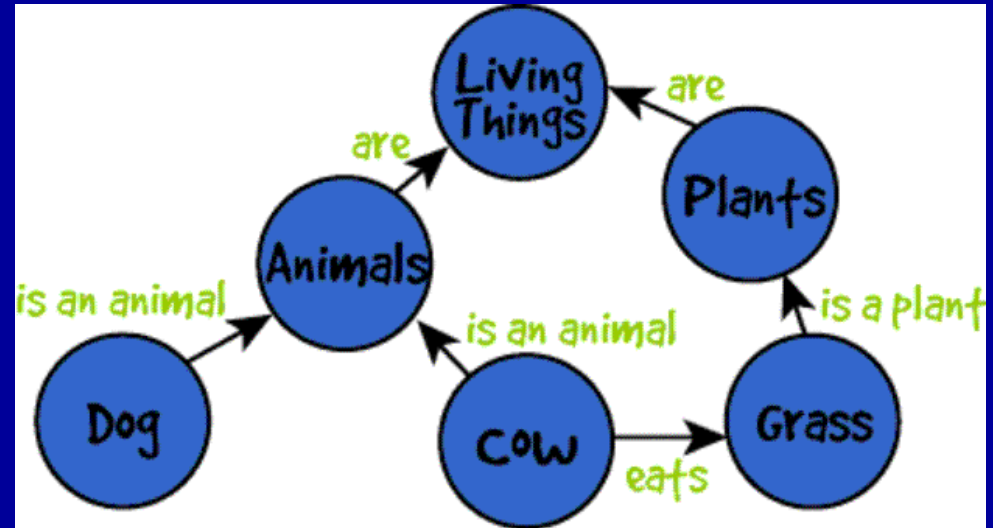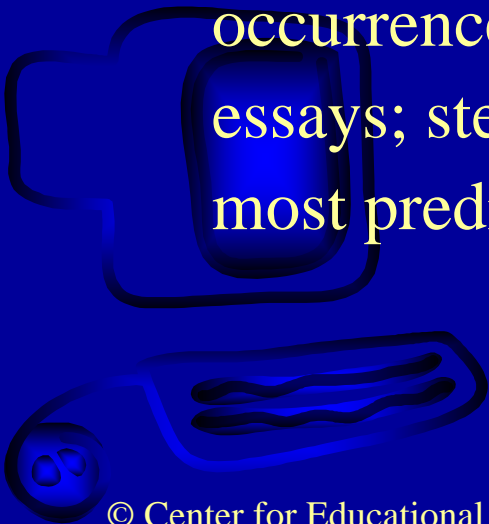
Concept Map (Klein, O'Neil, & Baker, 1998)

Image from http://www.graphic.org (2002)
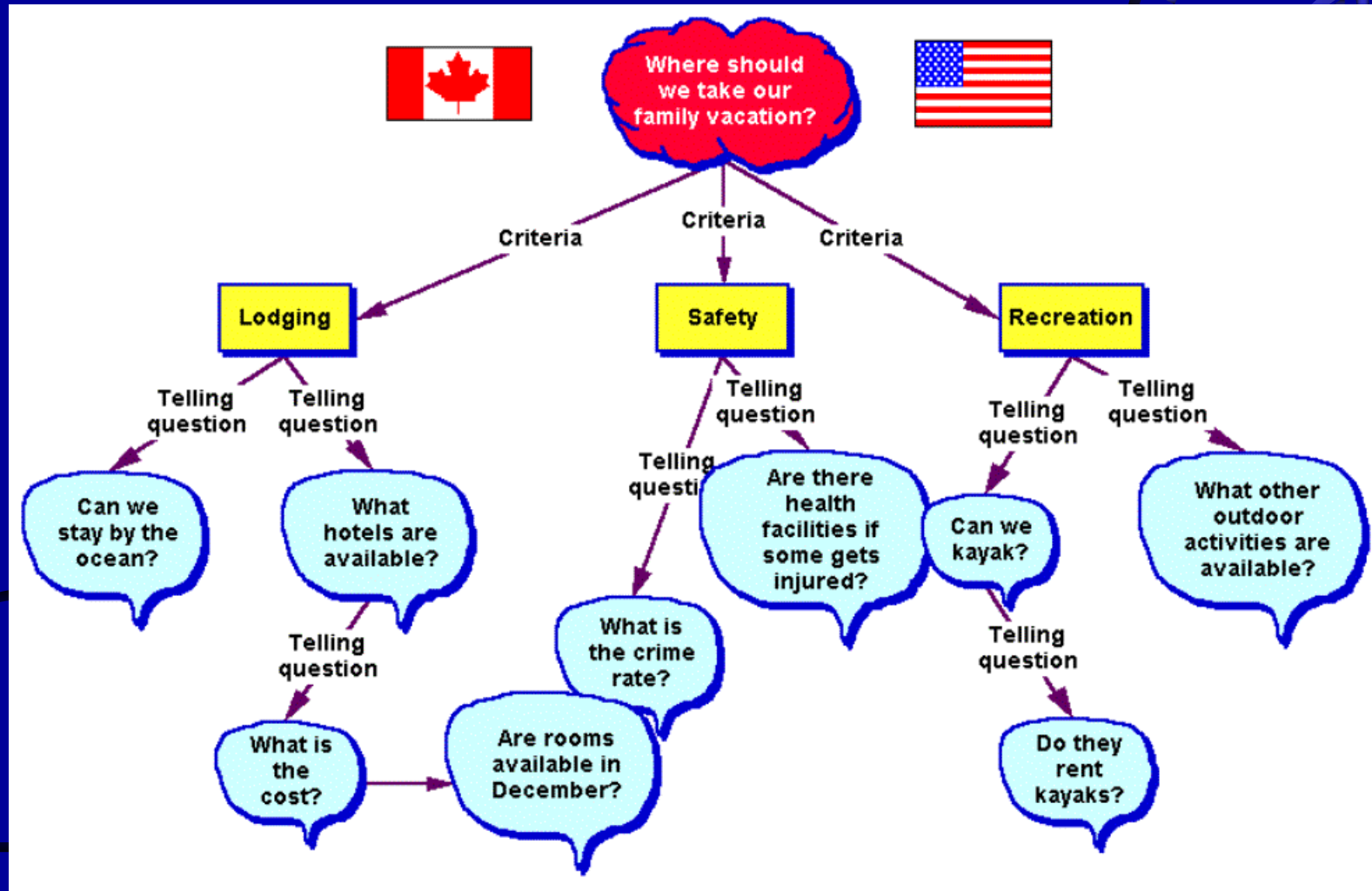
Automated Scoring of Text

- *Project Essay Grade* (Page, 1994; Page & Peterson, 1995)

  - Uses multiple regression to identify optimal set of surface features to predict score

- *E-rater* (Burstein, et al, 1998)

  - Implements computational linguistics to quantify occurrence of syntactical, topical, and rhetorical features in essays; stepwise regression then derives weights for features most predictive of scores by humans

Automated Scoring of Text (cont.)

- *Latent Semantic Analysis* (Landauer, Foltz, & Laham, 1998)
  - Relationships between words form semantic space
  - The LSA algorithm evolves a meaning for each word given semantic space between that word and every other word it is linked to

- *Text Categorization* (Larkey, 1998)
  - Linguistic expressions and contexts in texts used to classify them

Demo: PEG Automated Essay Scoring

Demo: LSA Automated Essay Scoring

Essay Submission Page

If you are using this site as part of a course, please fill in all the fields. If you are just testing the system, please just fill in the Name field

What is your name? April

What is your class codenumber?

Which class are you in?

How would you like your results outputted?
- Simple (grade only, takes about 5 seconds)
- Extended (grade and suggestions. **Note:-** this takes 20-30 seconds to complete, DEFAULT)

Describe McClelland and Rumelhart's Interactive Activation model of word recognition. How does it account for the Word Superiority Effect?

The interactive activation model of word recognition put forth by McClelland and Rumelhart is very interesting. It represents a significant step forward in the development of approaches to automated scoring of text, particularly in that it incorporates balloons and confetti. It includes assumptions about the underlying links between items and in the process of teasing out such links it promotes a syntactic network of lexical terms. Processing in this model takes place at three different levels: these are the subterranean, the sorta-underground, and the surface. While serial processing is optimal in many situations, this model focuses on the possibilities associated with parallel processing. The levels of the model interact at multiple levels as well -- some levels are effectively noninteractive except in cases of particularly unusual word structures while others directly influence each other in the context of processing. Activation within levels is bidirectional and is designed by McClelland and Rumelhart to faciliatate word recognition using all available information. Across levels, processing is structured as to as create an alphanumeric and syntactic network tht incorporates maximum flexibility for the task at hand. The word superiority effect is defined

Clear fields   Submit essay for grading

Demo: LSA Automated Essay Scoring

© Center for Educational Assessment, 2005

Free-Response Mathematics
(Braswell, 1990)

Item : 1

63.35
+ 8.58

Answer: [   ]

| ? Help | 📊 Graphs | **Next** | **Close** |

Image courtesy of ETS/C. Mills (2000)

Move Figures/Symbols into Pictograph
(Martinez, 1991; French & Godwin, 1996)

04:23 Remaining   Item 1 of 1

Choose one of the figures on the left and position it on the grid below to produce the graph of the function defined by $f(x) = x^2 - 1$.

Done

Image from Zenisky & Sireci (2002)

Move Figures/Symbols into Pictograph
(Martinez, 1991; French & Godwin, 1996)

Image from ETS/C. Mills (2000)

Graphical Modeling
(Bennett, Morley, & Quardt, 2000)

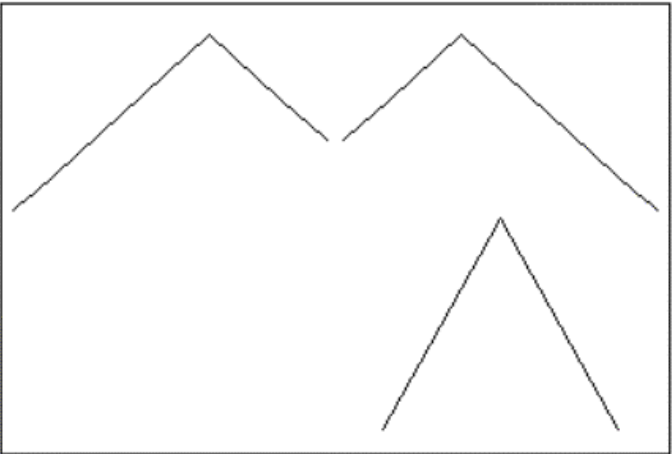Image from ETS/C. Mills (2000)

Graphical Modeling
(Bennett, Morley, & Quardt, 2000)

In the graph, shade a set of regions whose combined area is equal to

$$\sum_{n=1}^{3} f(n)$$

$y = f(x)$

Click on a region to shade it.

| Test | | | | | | | Help | Prev | Next |
| Quit | Time | Review | Mark | Erase | Calc | | | | |

Image courtesy of ETS/C. Mills (2000)

Multiple Numerical Response
(Mills, n.d.)

Image courtesy of C. Mills (2000)

Examples: Online Feedback/Reporting
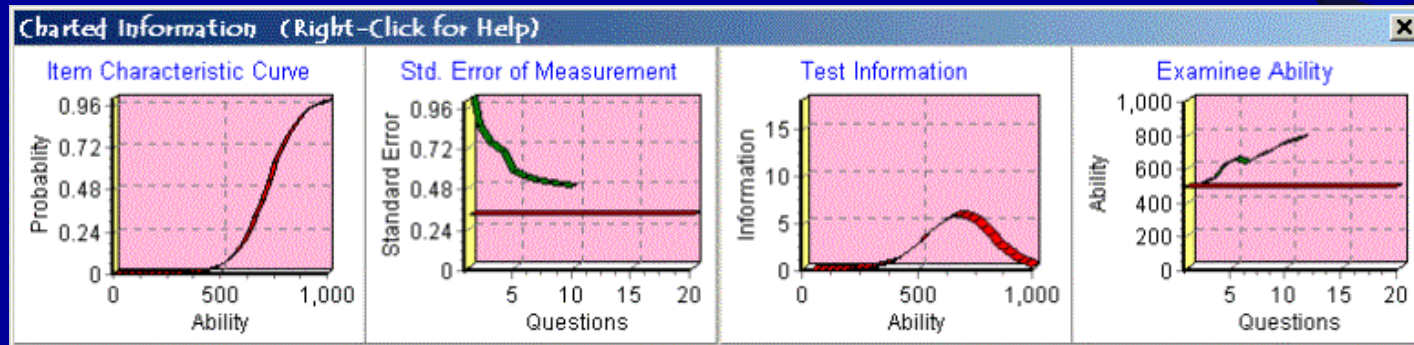
The example above illustrates some *measurement* feedback you might want back after taking a test…

At right is a sample score report from an adaptive math test found online

Score Report

Microsoft®
Where do you want to go today?®

Test Results:	Passed
Final Score:	760.80
Test Reliability:	90.01%
Total Questions:	10
Test Ended:	Your test ended because achieved a score significantly above the mastery level.

? Help   Graphs   ✔ Done
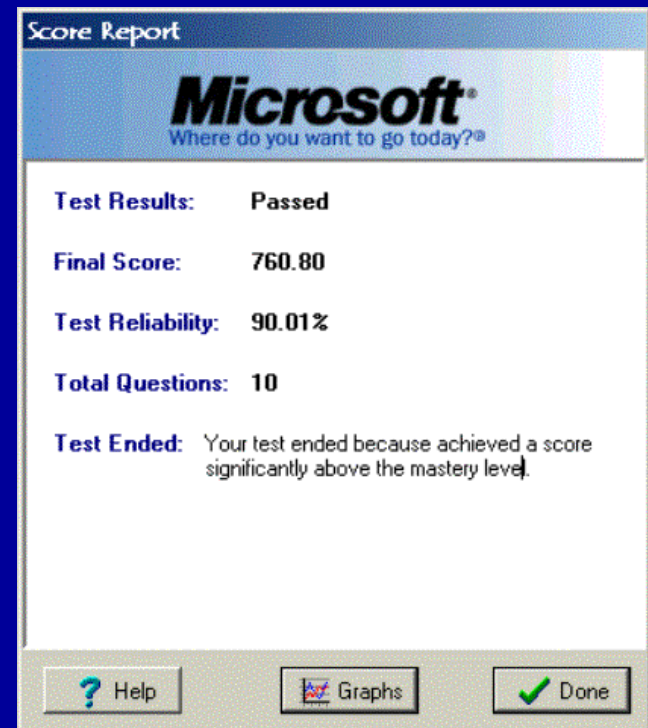
Computer-based essay
(Burstein, et al., 1998)

Analytical Writing Assessment - Analysis of an Argument

Read the argument and the instructions that follow it, and then make any notes that will help you plan your response. Begin typing your response in the box at the bottom of the screen.

Beginning

The following appeared as part of an article in a weekly newsmagazine.

"The country of Sacchar can best solve its current trade deficit problem by lowering the price of sugar, its primary export. Such an action would make Sacchar better able to compete for markets with other sugar-exporting countries. The sale of Sacchar's sugar abroad would increase, and this increase would substantially reduce Sacchar's trade deficit."

Discuss how well reasoned you find this argument. In your discussion be sure to

Cut
Paste
Undo

Test Quit | Section Exit | Time
? Help | Answer Confirm | Next

Image from ETS (2002)

Problem Solving Vignette/Simulation

Current Time: 2:34:32 PM

☐ Mark

You host a Web site for a customer named Just Togs. You want to change the default document so that users view Justtogs.htm instead of Default.htm when they connect by using the URL http://www.justtogs.com.

- Use Microsoft Management Console to change the default document to Justtogs.htm. (Click the **Microsoft Management Console** button.)

Microsoft Management Console

Click Next (or More) to continue.

Next   Previous   Help

Image courtesy of Microsoft (2000)