



# The ONPAR Project

Jim Bauman  
David Gabel  
Center for Applied Linguistics



The Center for Applied Linguistics is proud to be celebrating its 50th birthday in 2009. Please visit the CAL web site at [www.cal.org/cal50](http://www.cal.org/cal50) for some reflection on our past and plans for our future.

# Project Organization



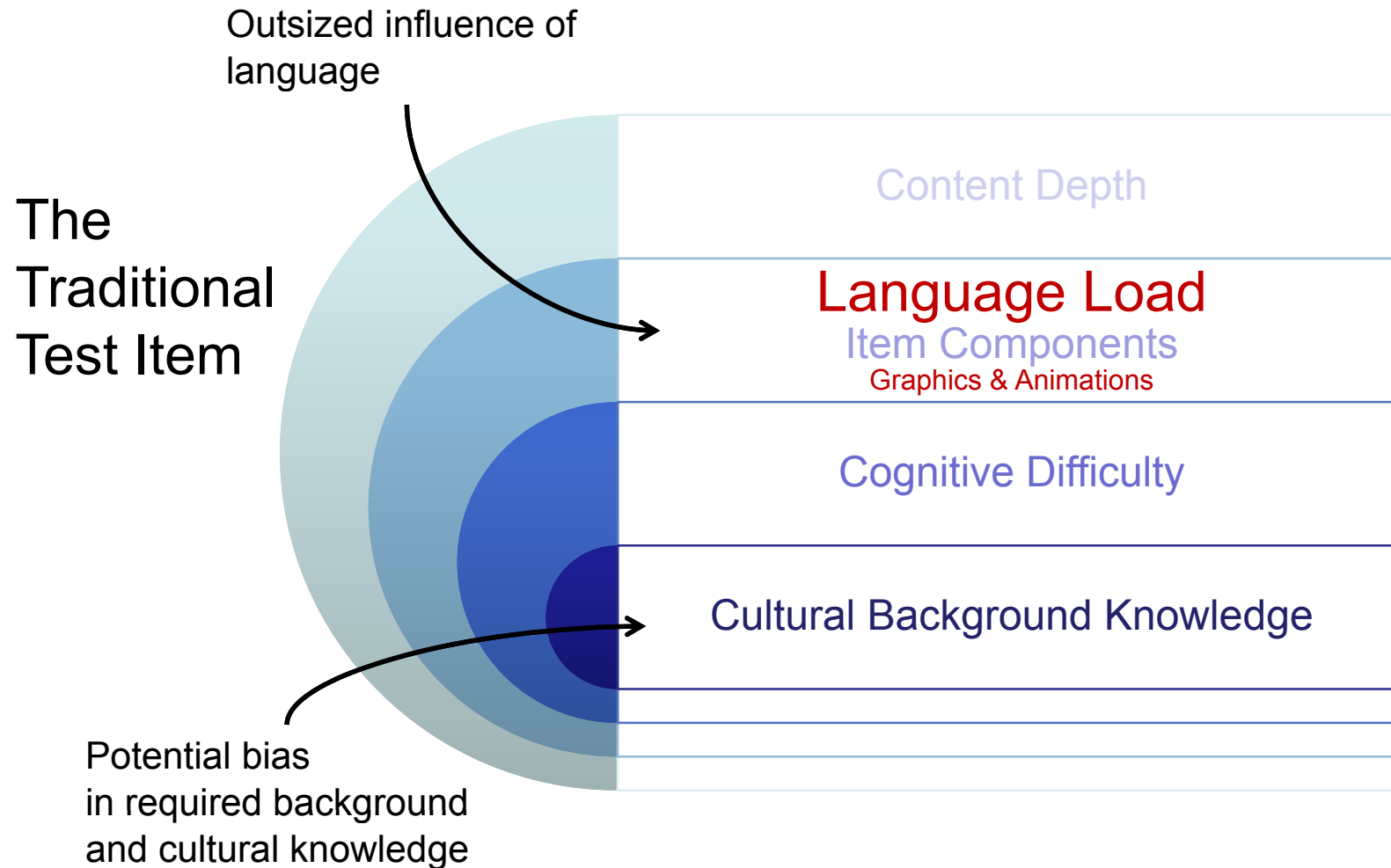
- Funded through two EAG's (Enhanced Assessment Grants) from the U.S. Department of Education
- Lead States
  - Science – Rhode Island
  - Math – Illinois
- Wisconsin Center for Educational Research via the WIDA Consortium
  - Tim Boals, Project Director
  - Rebecca Kopriva, Principal Investigator
- Center for Applied Linguistics
  - Jim Bauman, Program Manager
  - David Gabel, Lead Research Associate for Item Development
  - Cathy Cameron, Lead Research Associate for Research Operations
  - Laura Wright, Lead Research Associate for Linguistic Analysis

# Test Equity Issues for ELLs

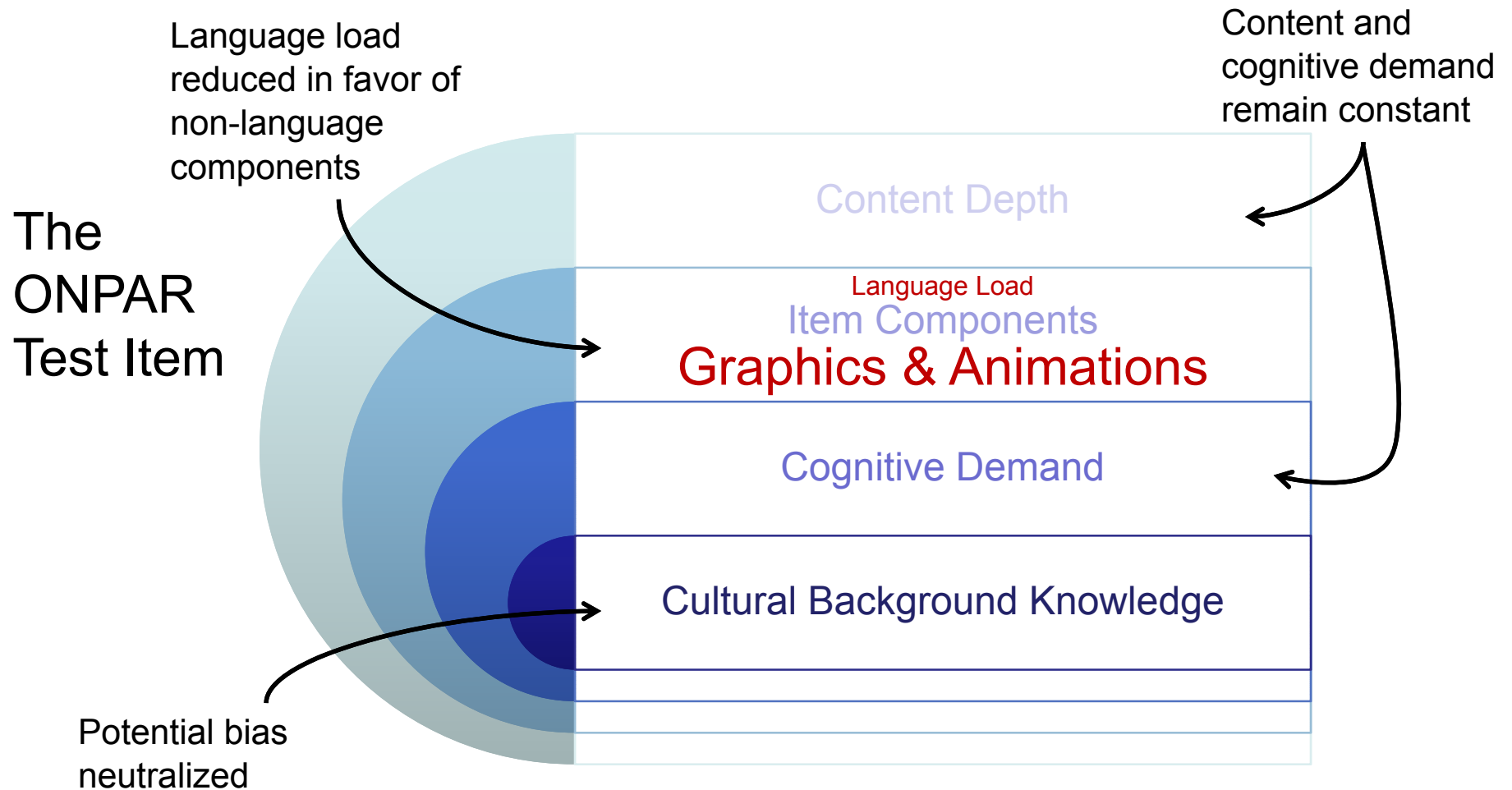
- Because of their lack of English language skill, ELLs do not perform well on:
  - Traditional paper-and-pencil based tests
    - Reading and writing demands exceed their capabilities
  - Multiple-choice items
    - Required discriminations between choices require fine-tuned language skills
  - Constructed-response
    - Productive language skills are lacking
  - Tests requiring cultural and background knowledge outside their experience
    - Unfamiliar contexts can confuse, rather than assist comprehension

How do you reliably assess science and math for ELLs who have real knowledge and skills in these areas?

# Equity Imbalance in Content Testing for ELLs



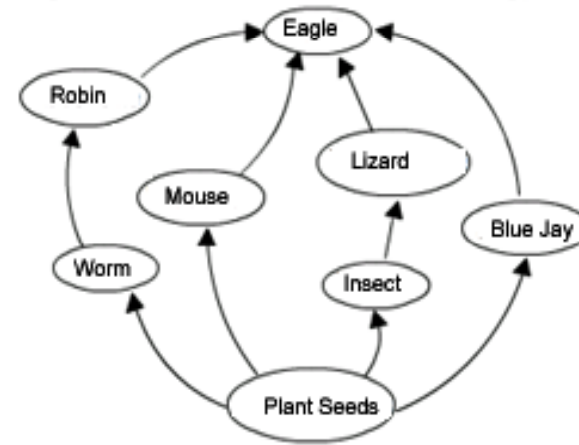
# Rebalancing Items for ELLs



# Exercise: Analyzing a Traditional Item

- Briefly describe what you think is the content target of this item.
- What are the cognitive demands of this item for the test taker. Describe the demands in terms of those specific to the content target and those irrelevant to the content target.
- What cultural and background knowledge assumptions does the item make?
- Describe your impressions of the language load of the item and the appropriateness of that load for English language learners at various proficiency levels in English?

The diagram below shows the flow of energy in a food web.



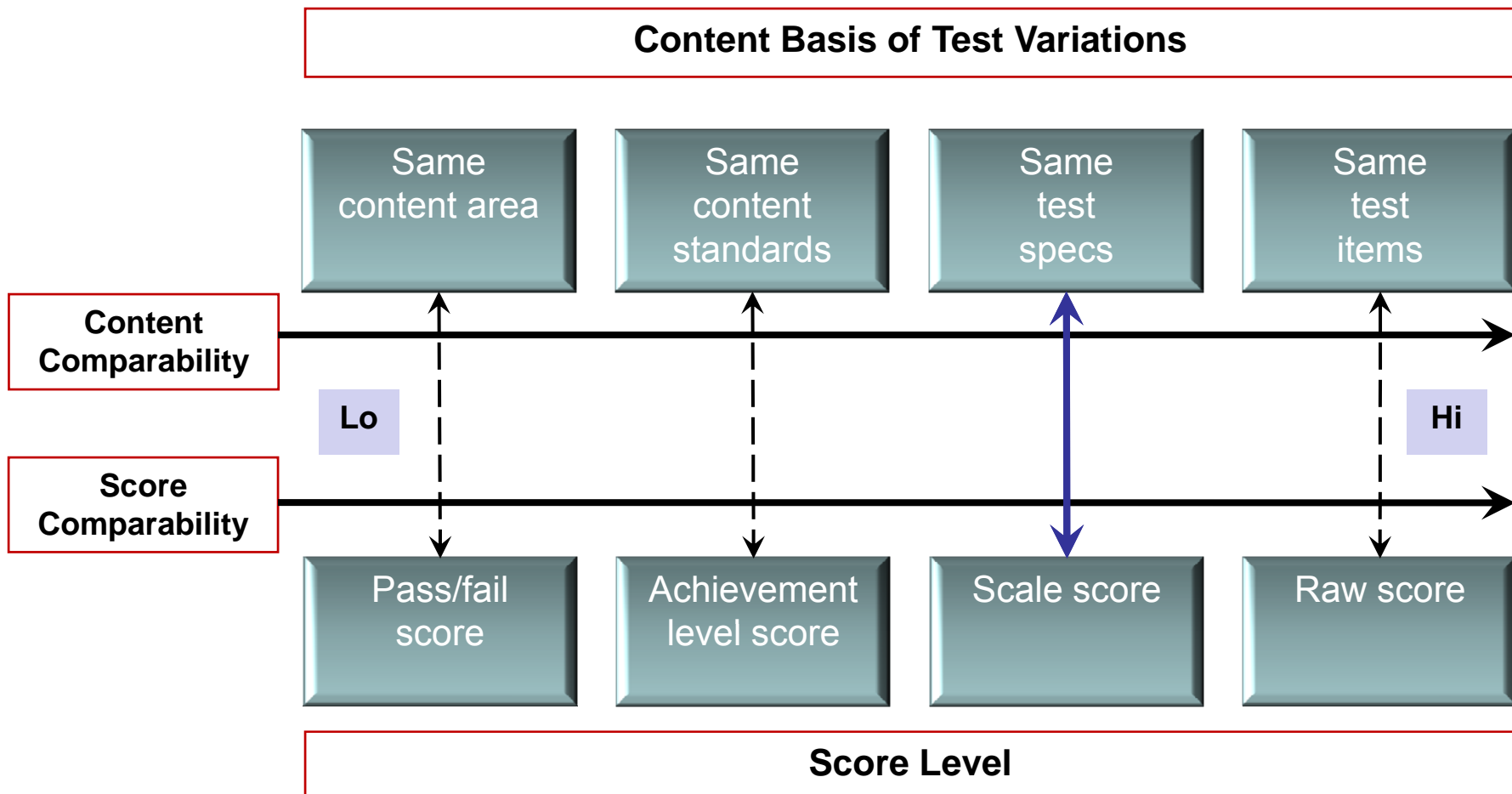
What is a relationship shown in this food web?

- A. a secondary consumer eats a secondary consumer
- B. a primary consumer eats a decomposer
- C. a decomposer eats a producer
- D. a producer eats a primary consumer

# ONPAR Project Objectives


- **Item development**
  - Create a core of items in science and math that successfully addresses a common set of standards at high depth of knowledge
  - Reduce the language load of these items without sacrificing content and cognitive complexity
  - Deemphasize multiple-choice item types in favor of item types more accessible to ELLs
- **Test development**
  - Supplement the core items with a set of items that reflect more local standards and lower depth of knowledge
  - Demonstrate comparability of test to traditional tests
- **Discourse framework development**
  - Situate the language used in the discourse genre of the test item
  - Build a discourse framework sensitive to text and non-text components
- **Computer-based testing**
  - Integrate distribution, scoring, and reporting functions within a computer-based test delivery platform
  - Develop automatic scoring algorithms for each item

# Continuum of Comparability



*Adapted from a model developed by  
Phoebe Winter*

# Demonstration Items

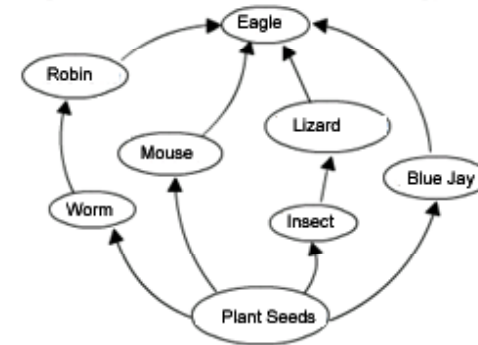
- Move the cursor to the icon for the item you want to view.
- If necessary, press **Ctrl-A** to change the cursor to the  symbol.
- Start your interaction with the test item.
- When you are finished, press the **Quit** button in the lower left corner of the navigation bar. You will return to your current slide in the presentation.

# Initial Design Decisions

- Deliver test items on computer
- Develop dynamic items using text, visuals, animations, and graphic organizers
- Develop a set of item types that permit interaction capable of demonstrating high depth of knowledge
- Support access as needed in a reduced language context

## Traditional Item

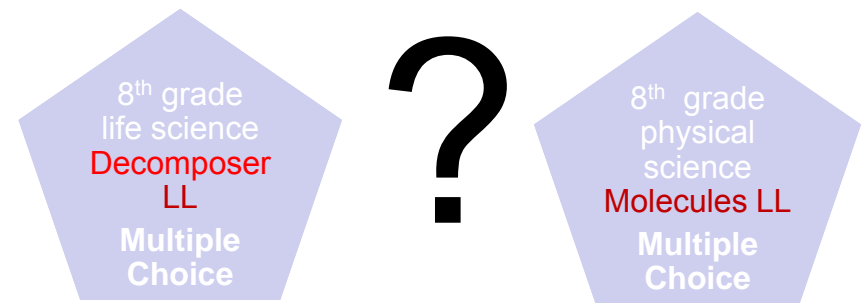
The diagram below shows the flow of energy in a food web.



What is a relationship shown in this food web?

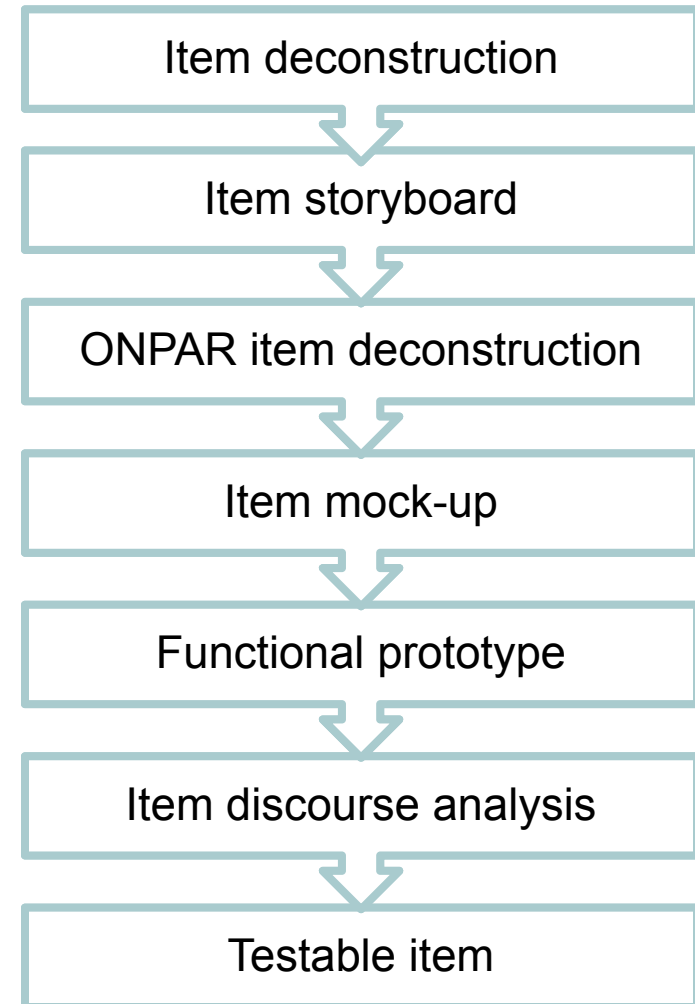
- A. a secondary consumer eats a secondary consumer
- B. a primary consumer eats a decomposer
- C. a decomposer eats a producer
- D. a producer eats a primary consumer

## ONPAR Item



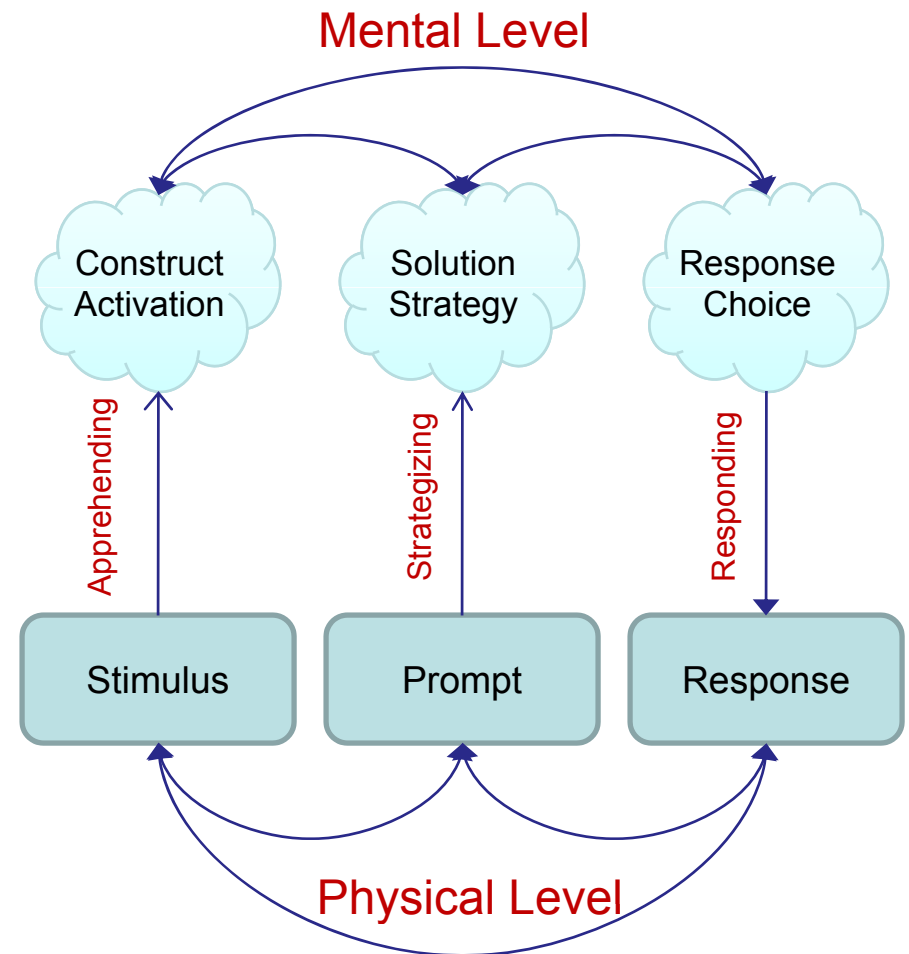
# Primary Stages in ONPAR Item Development

1. Detail construct-relevant and construct-irrelevant components of traditional test items.
2. Select appropriate item type to meet cognitive processing requirements for the item and conceptualize item content to match assessment target of traditional item.
3. Justify use of language and non-language components.
4. Create non-language components and lay out the item following standardization guidelines.
5. Program dynamic and interactive components.
6. Create low language and very low language versions of the item.
7. Ready item for trials.



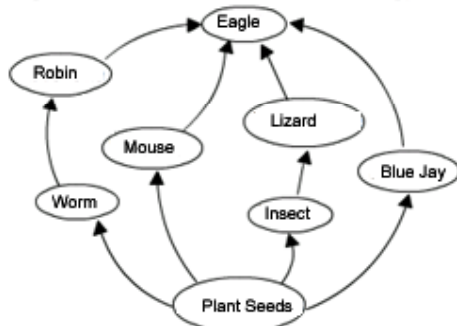
# Processing Model for a Test Item

- **Apprehending**
  - Does the student understand what the test item is generally about and what specific knowledge of the content he or she is being asked to provide?
- **Strategizing**
  - Is the student able to employ a strategy to solve the question, including being able to cognitively structure the asked for response?
- **Responding**
  - Is the student able to show an answer in a way that reliably demonstrates his or her true knowledge?



# Comparative Analysis of an Item

The diagram below shows the flow of energy in a food web.



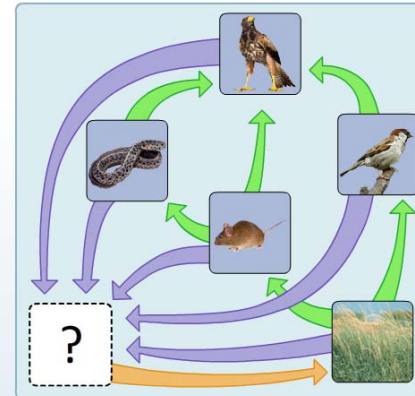
What is a relationship shown in this food web?

- A. a secondary consumer eats a secondary consumer
- B. a primary consumer eats a decomposer
- C. a decomposer eats a producer
- D. a producer eats a primary consumer

## General Test Item

1. **Construct:** illustrated food web; “flow of energy” is explicit
2. **Prompt:** “Relationships” → Food web
3. **Situated roles:** TG ↔ TT; Explicit question from TG
4. **Language use:** specialized academic sentence constructions, specialized vocabulary, English only
5. **Stimulus:** labeled diagram & text using specialized vocabulary
6. **Cohesion markers:** “Shows” → “shown”; “Food web” (“eats” (“consumers”, “producers”, “decomposers”))
7. **Item type:** traditional multiple-choice format; 1 option is correct

Complete the food web.



This ONPAR item is illustrated in its low language version.

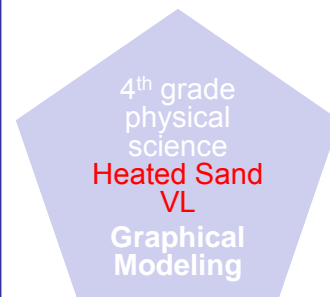
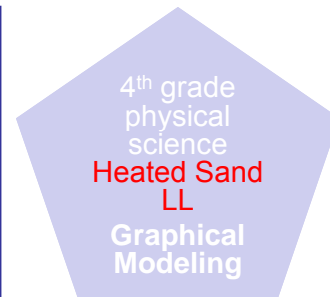
## ONPAR Item

1. **Construct:** illustrated food web; flow of energy is inferred
2. **Prompt:** rollover support on food web; gap in food web to indicate focus of task demand
3. **Situated roles:** TG ↔ TT; Explicit imperative from TG
4. **Language use:** conventional register, English with L1 support
5. **Stimulus:** illustrated diagram—no vocabulary, realistic images
6. **Cohesion markers:** Food web rollover; Select 1 icon → options → response space ← arrows within stimulus
7. **Item type:** select 1 cues response demand; Click-select option into bounded response space; 1 option is correct

Note: “TG” = the Test Giver; “TT” = the Test Taker; TG ↔ TT indicates an interaction between participants

## Low Language

- Unsupported language is aimed at test takers with mid level proficiency in English
- Prompts use simple, sentence level English.
- Prompts use L1 or English audio translations to accommodate comprehension.
- Links exist between many to most language components and their non-language counterparts.
- Level of inferencing required to comprehend the prompt is minimal.



## Very Low Language

- Unsupported language is aimed at test takers with low proficiency in English
- Prompts use simple, lexical or phrasal based English.
- Prompts are separated into a schema-referencing component and a task demand.
- No L1 translations are provided.
- Links exist between all language components and their non-language counterparts.
- Moderate level of inferencing is required.

# ONPAR Item Types—Part 1

ONPAR Item Types		
Item Type	Response Option	Description
<b>Multiple Choice</b>	Point and click	Alternatives represent a plausible set of options, only one of which will be correct.
<b>Extended Multiple Choice</b>	Point and click	Same as multiple choice, but with more than 5 and less than 20 possibilities. Could also involve selecting 2 choices from a list.
<b>Ordering</b>	Drag and drop	Arrange objects into a specified order.
<b>Matching</b>	Drag and drop	Match items in one column to items in another.
Select and Classify	Point and click; Drag and drop	Select words and/or images and arrange them into categories.
<b>Simple Relational Statement</b>	Point and click	Select a word or image that fits into a statement frame (grid). End result constitutes a well-formed proposition. Choices may be conditioned or unconditioned. Logical phrases (and/or/but cause/effect) could connect boxes.
<b>Representational Modeling</b>	Tool box; Connect points or objects	Use static or dynamic tools from a tool box to construct a response or complete a structured diagram.

*Bolded item types are currently in use in ONPAR items.*

# ONPAR Item Types—Part 2

## ONPAR Item Types (con't)

Item Type	Response Option	Description
<b>Graphical Modeling</b>	Point and click; Drag and drop; Slider bar	Model data by completing a graph (e.g., shading a portion of a geometric shape, drawing a line to represent trends in data, extending a bar graph).
<b>Stimulus Manipulation</b>	Visual highlighting; Slider bar; Menu bar	Manipulate an image or an animation in the item stimulus.
<b>Free/Open Response</b>	Key padding; Drawing	Could be a keypad, calculator response, or construction of mathematical sentences.
Complex Relational Statements	Drag and drop; Key padding	Complex type supported by a rich or extensible representation allowing for multiple means of engagement.
Problem Solving Vignette	All	A group of items that relate to each other and to a common stimulus with the goal of solving a problem.
<b>Thematic Probing</b>	All	A group of items that together involve a common or closely related stimulus. The item parts may or may not be dependent on one other.
<b>Schema Completion</b>	Drag and drop	Complete a schematic diagram by selecting appropriate options from a group of choices

*Bolded item types are currently in use in ONPAR items.*

# Item Standardization Features

- Text and Font properties
- Icons
- Response options
- Item stimulus
- Response space
- Navigation buttons

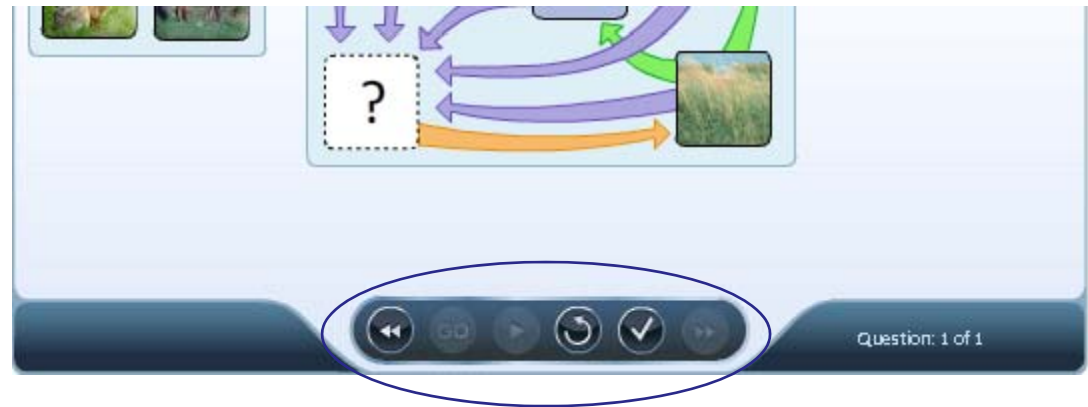
The screenshot displays a digital assessment interface for a food web task. At the top, a speaker icon and the text "Complete the food web." are visible. Below this is a green "Select 1" button with a hand cursor icon. A 3x2 grid of response options includes a lion, mushrooms, an ant, a tree, a rabbit, and a cow. The main stimulus is a food web diagram with a central mouse, a snake, a hawk, a sparrow, and a field of grass. A dashed box with a question mark is positioned at the bottom left of the diagram. Navigation buttons (back, go, forward, refresh, check, next) are located at the bottom of the interface, along with the text "Question: 1 of 1".

# Navigation Bar



## Navigation Bar

- Includes buttons to initiate an item, replay animations, reset the response space, submit items, and move between parts of items that contain multiple parts.
- Bar also contains question tracking indicator.
- The bar operates the same in both low and very low language versions.
- Grayed out buttons are not available in the current context.



# Text and Font Properties

Return to Previous Screen

## Text Properties

- Low language version uses full sentences in prompt.
- Very low language version uses word or phrase structures in prompt. Cognitive demand is separated from the schema activator by a vertical line.
- Prompt is located in top left corner of item.

## Font Properties

- Same type face and size are used consistently for all text.
- Words underlined in blue control roll-overs or pop-ups that clue the meaning of the words



Low language version



Very Low language version



# Item Icons



## Speaker Icon

- Activates an English or L1 audio of item prompt
- Only in low language versions



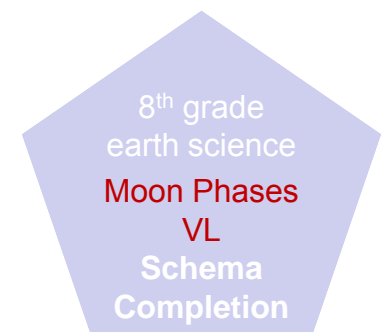
Low language version

## Item Type Icon

- Identifies the item's response mechanism
- Animated to demonstrate response procedure
- Identical in the low and very low language versions
- The same item type may be expressed with different icons, depending on the required type of physical response



Very Low language version

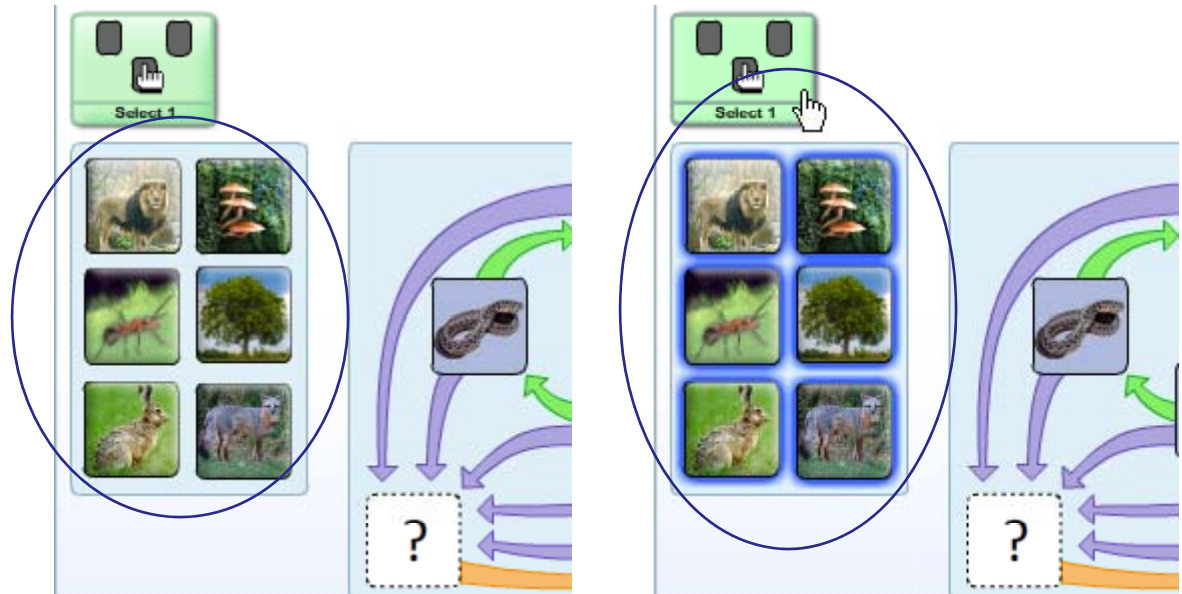


# Response Options

Return to Previous Screen

## Response Options

- The look, number, and position of response options varies with the item type (and icon).
- In the Select item type illustrated, the options are represented by a set of illustrations, one of which is the correct or key value.
- A roll-over on the icon highlights and identifies the location of the response options.



Low language  
and very low  
language version

Response options  
highlighted by roll-  
over on icon

4<sup>th</sup> grade  
physical science  
**Buoyancy LL**  
Stimulus  
Manipulation

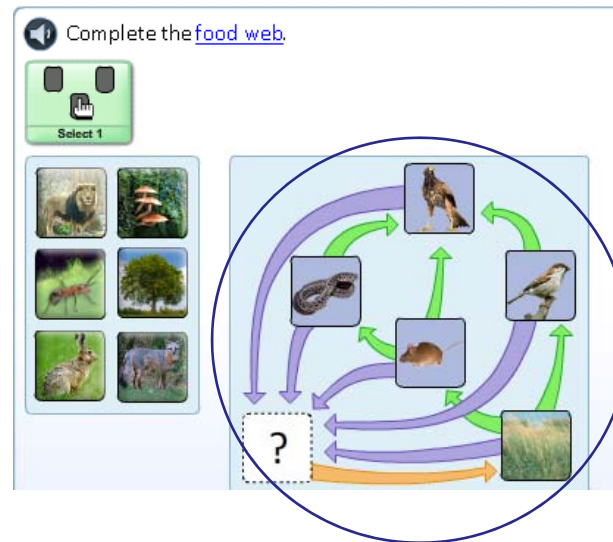
8<sup>th</sup> grade  
physical science  
**Circuit Diagram**  
LL  
Representational Modeling

# Item Stimulus

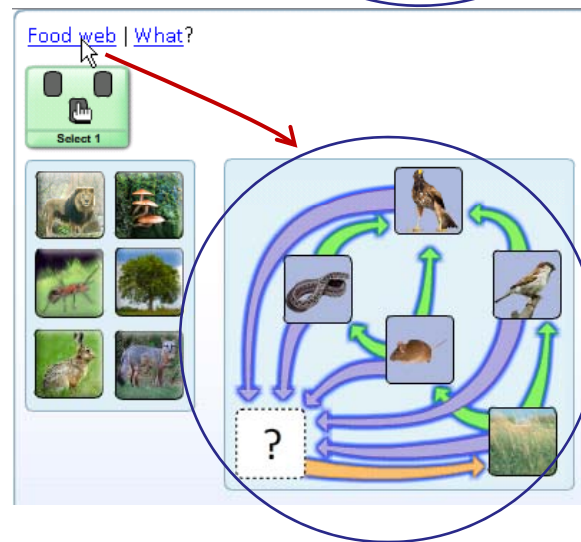
Return to Previous Screen

## Item Stimulus

- The item stimulus provides the situational context for the item's content.
- The structure of the stimulus varies with the item type. In many instances the stimulus is animated.
- The response space can be structurally included in the stimulus (as illustrated here).
- The stimulus occupies the center right of the item and is structured the same for both low and very low language versions.
- The stimulus is highlighted by a roll-over on a key word in the question prompt.



Low language version



Very low language version  
Stimulus highlights at roll-over on the cue word in the prompt.

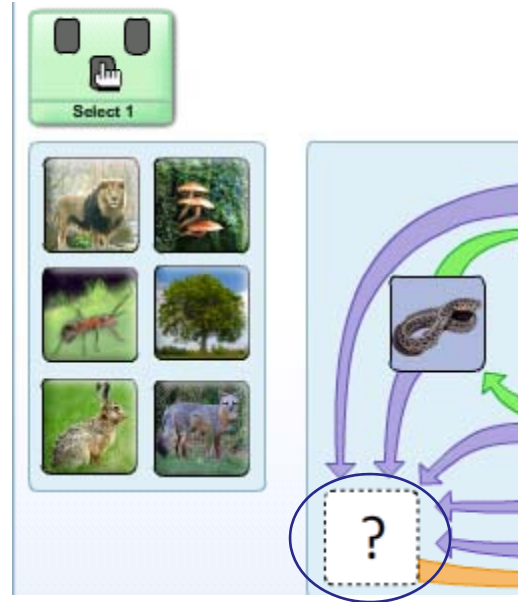
8th grade  
physical  
science  
**Force Box VL**  
Simple  
Relational  
Statement

# Response Space

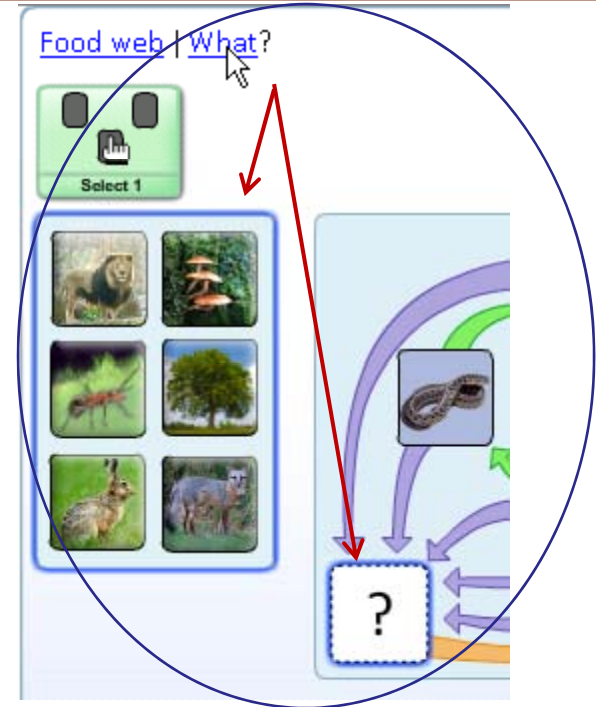
Return to Previous Screen

## Response Space

- The structure of the response space varies with the item type.
- The space is typically located within or below the stimulus area.
- The response space is typically cued with a “?” bounded within a dotted line frame.



Low language version



Very low language version

Rollover on the question prompt highlights both response options and response space

4<sup>th</sup> grade  
life science  
**Pond Ecosystem**  
LL  
Matching &  
Relational  
Statement

4<sup>th</sup> grade  
life science  
**Food Web**  
**Crisis LL**  
Graphical  
Modeling

# Item Validation Procedures

- Selected items are presented individually to students in cognitive labs.
- Items are assembled into the testing platform.
- Student access questionnaire is developed for selected items.
- Items are administered to groups of students in controlled trials.
- Items are analyzed for their behavior with the various groups.
- Items are judged by an expert panel for content coverage and for comparability to their traditional item models.
- Successful items are assembled into operational forms and field tested.
- Field test results are submitted to item analysis.

# Research Agenda: ONPAR Science Grant & Math Grant

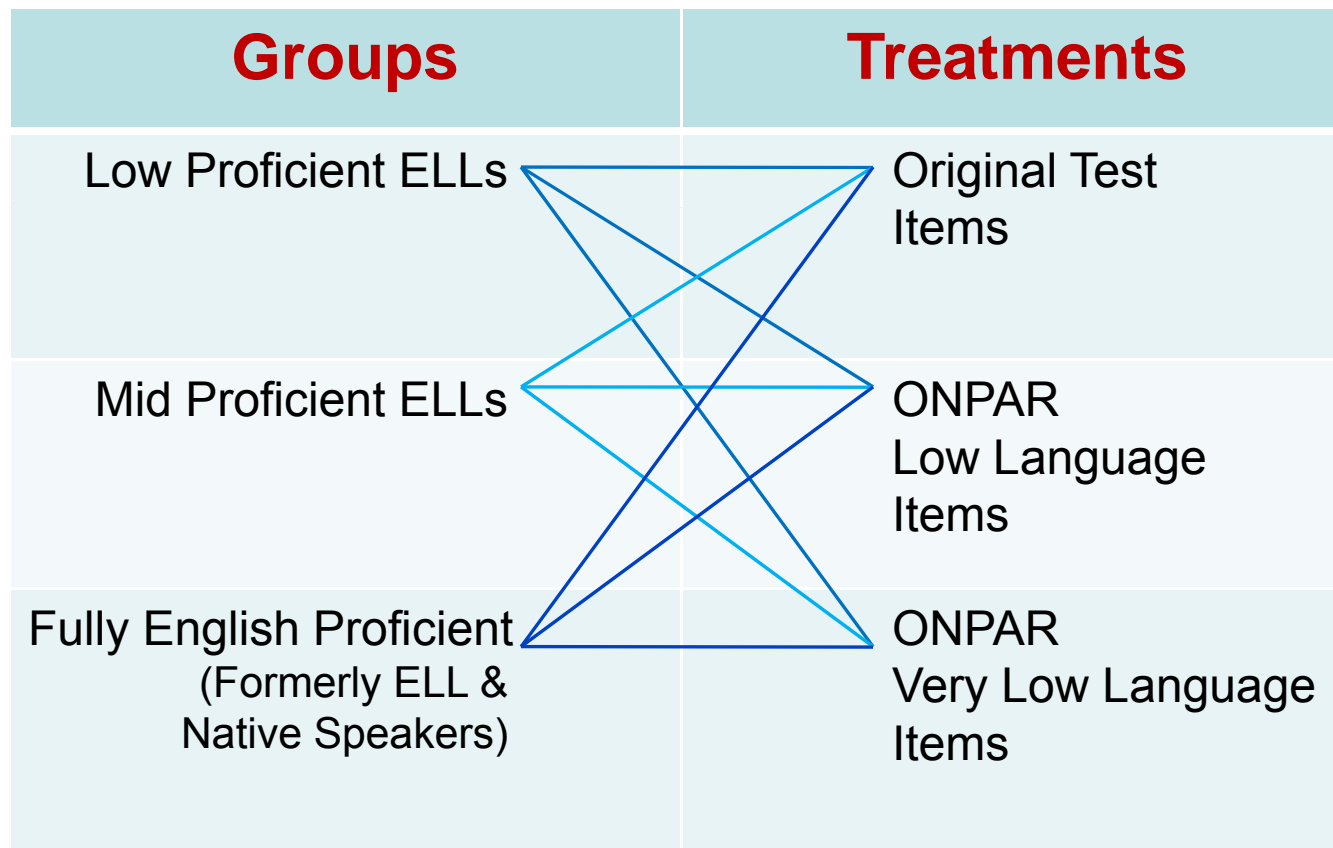


- Cognitive labs with:
  - ELLs at different levels of proficiency
  - Native English speakers
- Independent judgment review of cognitive demands in both ONPAR and standard items to verify:
  - Science and math targets
  - Cognitive demands in ONPAR and standard items
  - Facilitative and/or inhibitory effects of construct-irrelevant item components
- Controlled studies with science items using ONPAR and standard items
- Large-scale study with items to investigate convergent and discriminant validity vs. other salient indicators

## Timeline

- Cog Labs
  - Computer-based testing issues—Spring 2008
  - Cognitive-Linguistic issues—Spring 2009
- Judgment Reviews
  - Science—Jan 2009
  - Math—Jan 2010
- Controlled Studies
  - Science 8<sup>th</sup> Grade—Spring 2008 & Fall 2008
  - Science 4<sup>th</sup> Grade—Fall 2008
- Large-Scale Trials
  - Math & Science Fall 2009

# ONPAR Research Design



# The Issue of Item Accessibility

- How do ELLs know what the item is about?  
How do they know how they are expected to respond?
  - How does the mix of language and non-language text components influence accessibility?
  - How do students tackle the inferencing burden of the item? (Note the inferencing burden is higher for very low language items.)
  - What influence does the item type have in promoting or hindering accessibility?
  - How does the mix of and relations between construct-relevant and construct-irrelevant components affect accessibility?
  - What kinds of home language (L1) supports affect accessibility?

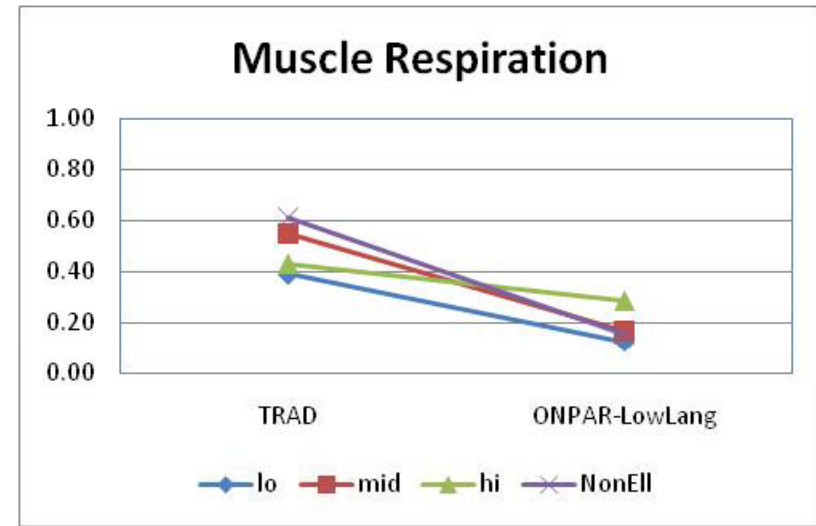
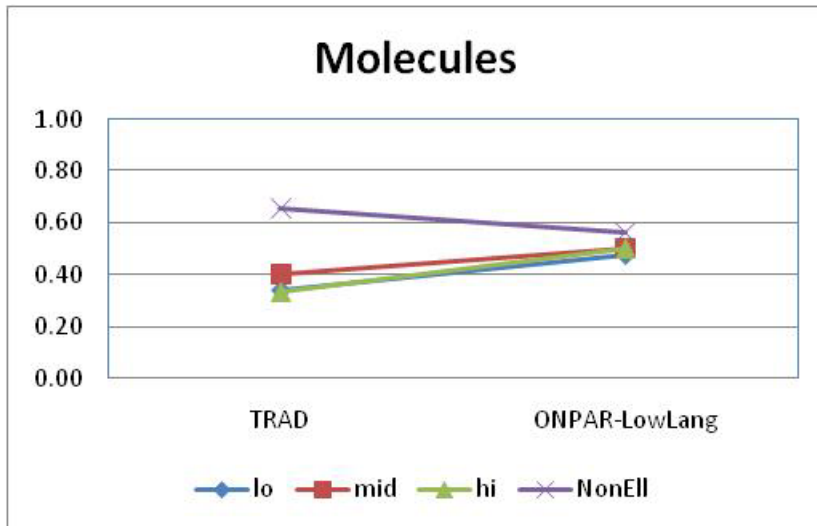


4<sup>th</sup> grade  
life science  
Rabbit  
Adaptation LL  
Matching

# Disentangling Accessibility from Ability

- A measure of student's access to particular items was constructed using a questionnaire test takers responded to following administration of the test itself.
  - Measure was based on two nested sets of questions:
    - To identify the assessment target and account for any reason (ability and/or accessibility) that made the identification impossible or difficult.
    - To indicate if they were able to give their preferred answer and account for any reason (ability and/or accessibility) that made the answer production impossible or difficult.
  - The scale ranged from 0 to 16, with 0 indicative of the lowest rating of accessibility.
- The ability measure was derived from teacher ratings of a student's proficiency in each of the item's assessment targets.
  - Scores ranged from 0 to 13 (the number of items tested)
  - Students were rated High if they scored above the mean and Low if below the mean.

# Item Performance



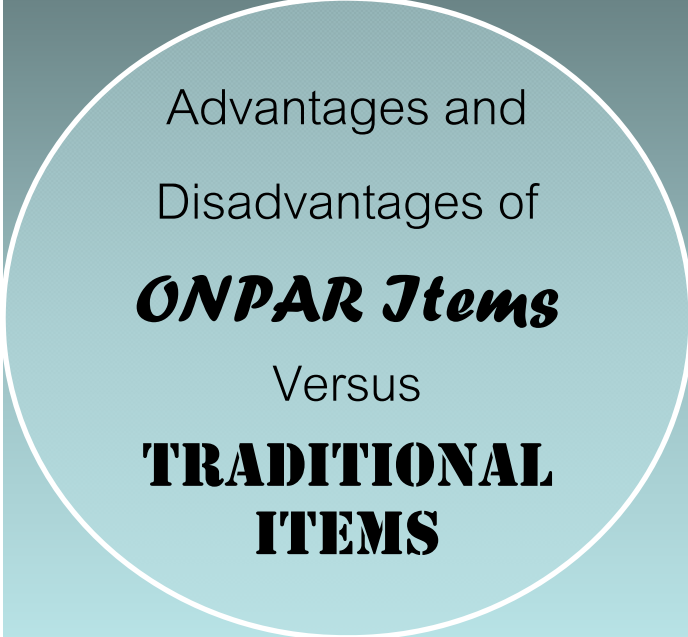
8<sup>th</sup> grade  
physical  
science  
**Molecules LL**  
Multiple  
Choice

- These two charts show items in which there is convergence of scores on the ONPAR form – Low language version. This is the desired result.
- The Molecules item (8<sup>th</sup> grade) is a multiple choice item that has a similar number of options on both the Traditional and ONPAR forms. Consequently, the guessing parameter is identical in both forms.
- The Muscle Respiration item (8<sup>th</sup> grade) has a very different guessing parameter on the two forms.
  - Chance on Traditional: 1/4
  - Chance on ONPAR:  $1/6 * 1/81 = 1/486$  (approximately 0)

8<sup>th</sup> grade  
life science  
**Muscle  
Respiration LL**  
Graphical  
Modeling

# Advantages & Disadvantages to ONPAR Approach

- Advantages
  - Creates more equitable items for ELLs (and perhaps other special populations)
  - Aligns better with inquiry-based and interactive instructional approaches
  - Offers possibility of embedding and integrating accommodations for test takers
  - More motivating for test takers
- Disadvantages
  - Cost of item construction is higher than for traditional items
  - Assessment targets focused specifically on academic language are not treatable
  - Requires technical infrastructure for testing which is still not universally available



Advantages and  
Disadvantages of  
***ONPAR Items***  
Versus  
**TRADITIONAL  
ITEMS**

- SEA involvement
  - Feasibility of computer-based testing of ELLs
  - Aligning ONPAR with state content tests
  - Activating the core-supplement strategy to build an operational test
  - Assembling and analyzing test data to support validation arguments
  - Motivating and supporting districts
  - Field testing with populations other than ELLs
- LEA involvement
  - Recruitment of schools and arranging for testing facilities
  - Training & preparation of test administrators
  - Professional development for teachers



For updated information  
about ONPAR, please visit  
us *after* November 1, 2008  
at  
[www.cal.org/onpar](http://www.cal.org/onpar)