

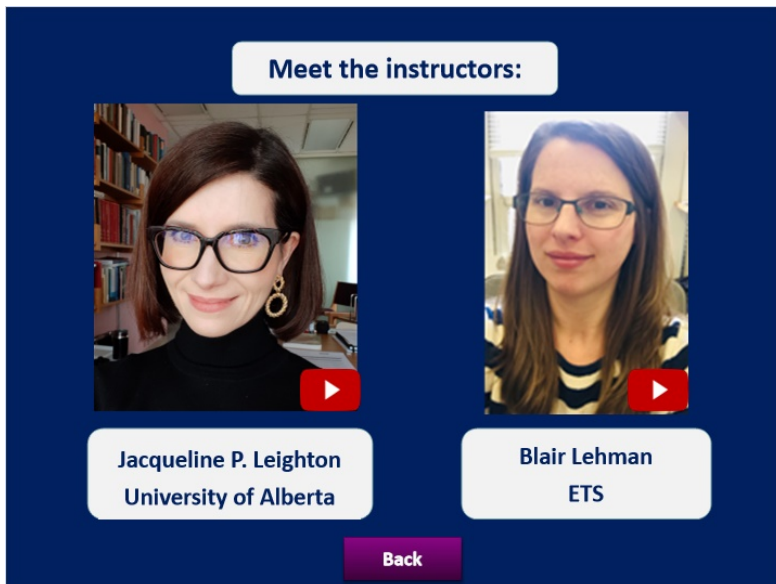
# DM12 SLIDES (Verbal Reports, Version 1.1)

## 1. Module Overview

### 1.1 Module Cover



### 1.2 Instructors




### 1.3 Designers

Meet the designers:




Xi Lu  
FSU



André A. Rupp  
ETS

Special Thanks:   

### 1.4 Welcome



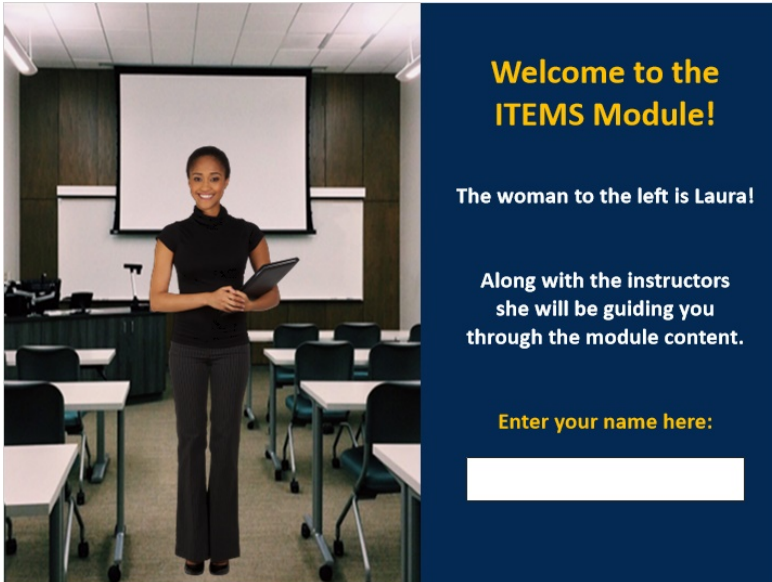
**Welcome to the  
ITEMS Module!**

The woman to the left is Laura!

Along with the instructors  
she will be guiding you  
through the module content.

Enter your name here:

## Untitled Layer 1 (Slide Layer)



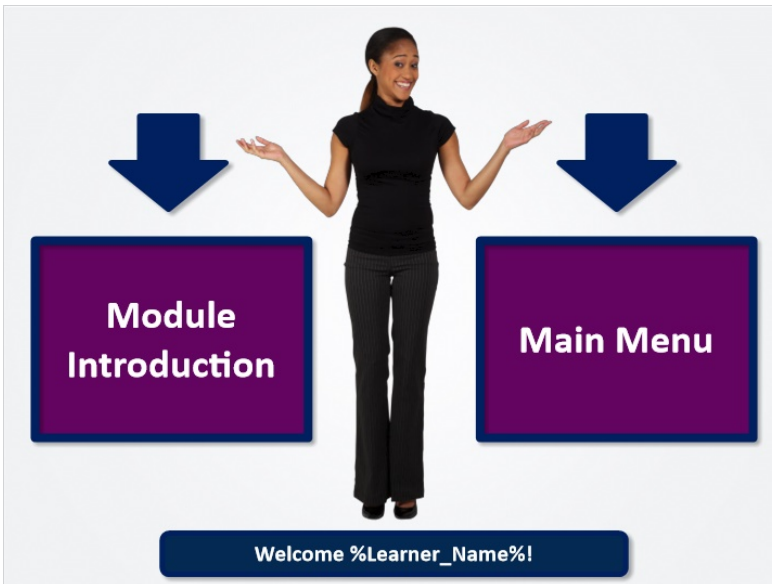
**Welcome to the  
ITEMS Module!**

The woman to the left is Laura!

Along with the instructors  
she will be guiding you  
through the module content.

Enter your name here:

### 1.5 Path Choice



**Module  
Introduction**

**Main Menu**

Welcome %Learner\_Name%!

## 1.6 Main Menu

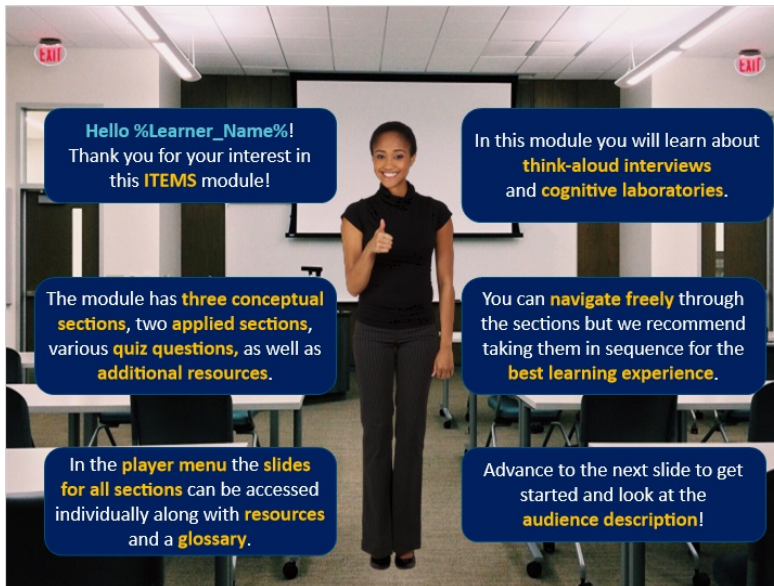


## Navigation (Slide Layer)





## 1.7 Overview




## 1.8 Target Audience

### Target Audience

Anyone who would like a gentle but technically sound introduction to this topic:

- Pre-service teacher preparation educators with a research mindset
- Graduate students in psychometrics, teacher education or educational leadership
- PreK-12 classroom teachers, teacher groups, teacher leaders
- School administrators responsible for teacher evaluation
- School/district assessment and curriculum leaders/facilitators
- Professional developers for in-service teachers
- Educational researchers and test designers



We hope that the module finds other meaningful uses as well no matter what your role in an organization is!

## 1.9 Prerequisites

### Five Basic Prerequisites

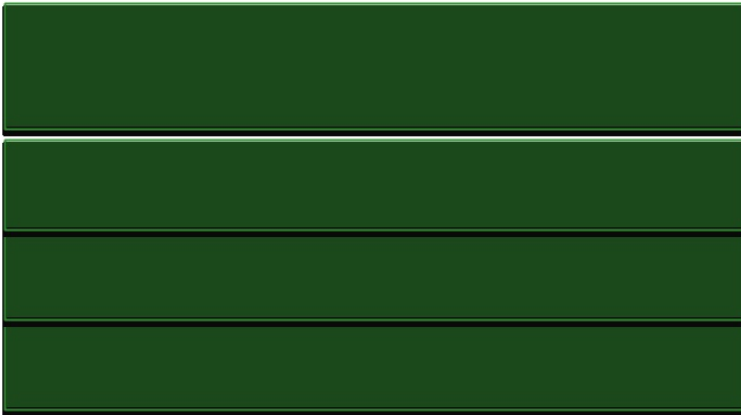
Learners should have at least a working knowledge of the following content areas:

- Human ethical protocols
- Basic psychological perspectives of human learning
- Basic tenets of constructivist and information-processing theory
- Key ideas in the *Standards for Educational and Psychological Testing*
- Applied research design and simple descriptive and inferential statistics

Resources

## Resources (Slide Layer)

### References



Click on each citation to be brought to the publisher web page

Back

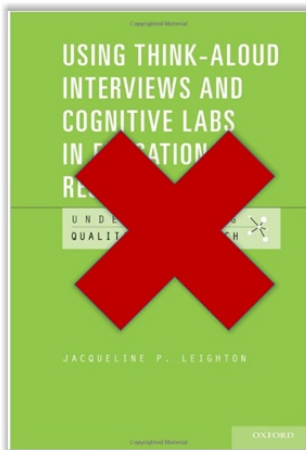
## 1.10 Expectations (I)



Let's discuss expectations....

## 1.11 Expectations (II)

### ITEMS Modules in Context




## 1.12 Resources

**Resources**

**Module Citation**


Leighton, J. P., & Lehman, B. (2020). Think-aloud interviews and cognitive labs (Digital ITEMS Module 12). *Educational Measurement: Issues and Practice*, 39(1), XX-XX.



**Additional Resources**


## References (Slide Layer)

**Resources**



Leighton, J. P. (2017). *Using think-aloud interviews and cognitive labs in educational research*. New York, NY: Oxford University Press.

Willis, G. B. (2015). *Analysis of the cognitive interview in questionnaire design*. New York, NY: Oxford University Press.



**Back**

### 1.13 Objectives

**Learning Objectives**

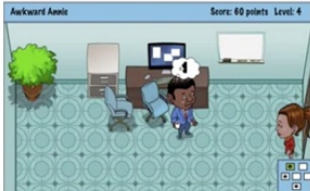
1. **Differentiate** between historical events and theoretical concepts underlying think aloud methods and cognitive labs
2. **Differentiate** methods and procedures for think-aloud interviews and cognitive labs along with the research situations where each applies
3. **Use** computer-based tools such as the ABC tool to collect verbal reports and understand how this differs from traditional methods
4. **Understand** differences in the analysis of verbal reports using pre-existing cognitive models or thematic analyses
5. **Evaluate** evidence required and collected for substantiating claims about specific forms of response processes
6. **Understand** the importance of inter-rater agreement indices for interpreting verbal reports

### 1.14 Copyright Statements

**Copyright Notice**

In the **Practice section** of this module, the instructors use the **ABC tool** as well as videos from a **game prototype** called **Awkward Annie**.

The tool and all game content as well as the **concept** of the game are protected by **ETS copyright**.



Copyright © 2020 by Educational Testing Service (ETS). All rights reserved.



## 2. Section 1: Conceptual Foundations

### 2.1 Cover: Section 1

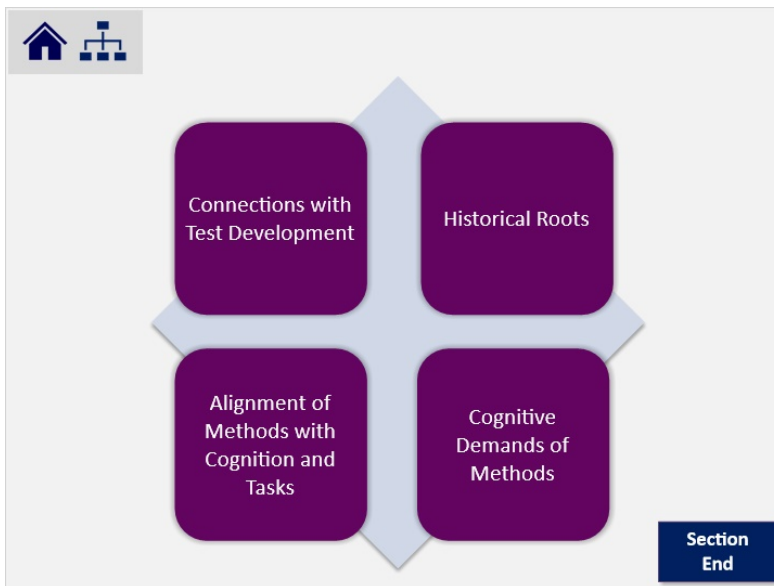


### 2.2 Objectives: Section 1

**Learning Objectives**

1. Describe professional standards associated with using think alouds and cognitive labs
2. Describe the different historical foundations for the two methods
3. Analyze task features to match them up with the appropriate methods
4. Apply information-processing evidence in the selection of appropriate methods for measuring unobservable knowledge and skills



## 2.3 Topic Selection



## 2.4 Bookmark: Standards



## 2.5 Standards I





### Evidence about Response Processes

**Standard 1.12**

If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument of validity, similar information should be provided.

(AERA, APA, & NCME, 2014, p. 26)

## 2.6 Standards II





### Evidence about Response Processes

Evidence based on response processes generally comes from analyses of individual responses...While evidence about response processes may be central in settings where explicit claims about response processes are made by test developers or where inferences about responses are made by test users, there are many other cases where claims about response processes are not part of the validity argument.


(AERA, APA, & NCME, 2014, p. 15)

## 2.7 Standards III






### Verbal Reports

- ✓ Response processes are **unobservable**
- ✓ Methods are needed to create **behavioral indicators** of these unobservable processes
- ✓ **Think-alouds** and **cognitive laboratories ("cog labs")** are key methods to collect evidence of response processes
- ✓ The data collected are called **verbal reports**



## 2.8 Bookend: Professional Standards

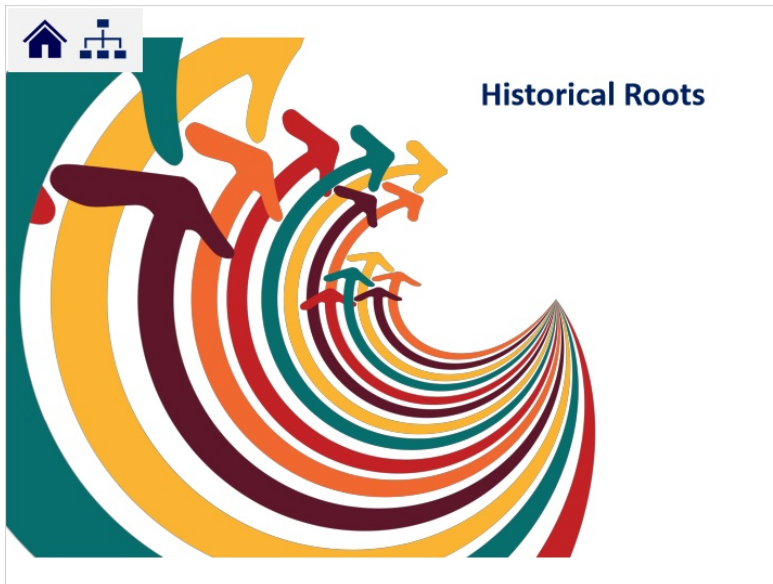






This is the end of this part.

Topics

## 2.9 Bookmark: Historical Roots



## 2.10 Origins of Methods





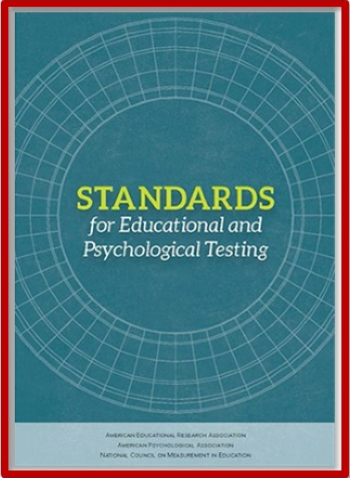
### Origins of Methods

- Although both **think aloud interviews** and **cognitive labs** involve the collection of **verbal reports** from participants, each interview method originated in response to distinct **research goals**.
- Cognitive labs or interviews emerged as a method to evaluate respondents' **comprehension of survey questions** and to **minimize social desirability**. This led to a **movement** in 1980s known as **Cognitive Aspects of Survey Methodology (CASM)**.



## 2.11 Untitled Slide










The Standards are a product of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

Published collaboratively by the three organizations since 1966, this manual represents the gold standard in the guidance of testing in the United States, Canada and many other countries.

The current edition is from 2014.  
Click on the images for more information.





## 2.12 Orgins of Cog Labs





### Cognitive Laboratories

Cognitive labs or interviews emerged as a method to evaluate respondents' **comprehension of survey questions** and to **minimize social desirability**.

Lots of **conversation, probing, and elaboration** to find out whether items were comprehensible in terms of **wording and format**.




## 2.13 Origins of Think-alouds





### Think-aloud Interviews

Think-aloud interviews emerged as a method to **validate information-processing models** of human problem-solving  
(Ericsson & Simon, 1980,1993)



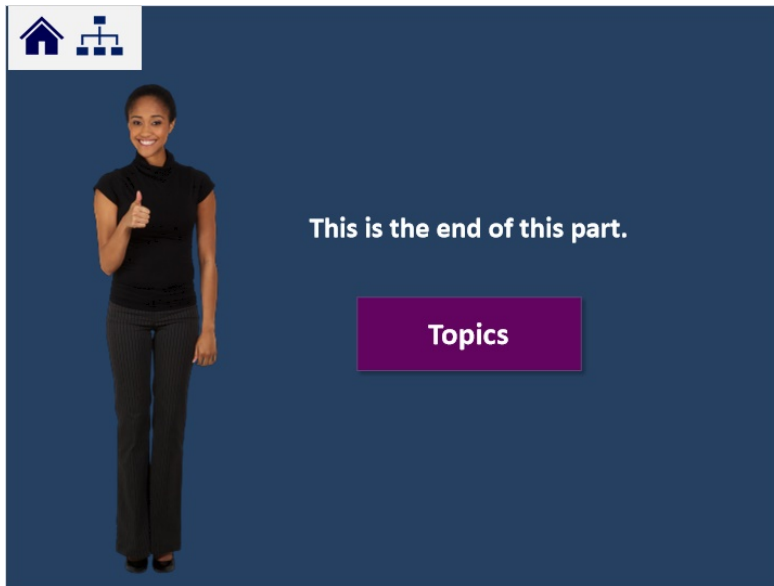
## 2.14 Origins II



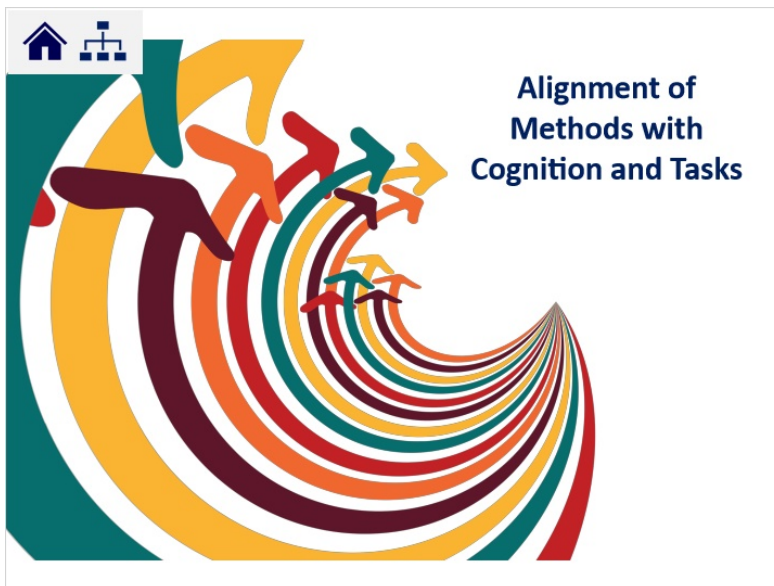
### Implications

- One perspective is that these distinctions are **purely historical** and **semantical / terminological**.
- It matters insofar as each is linked to **differences in rigor** for collecting evidence about **different types of cognitive processes**.
- It is critical that investigators **fully disclose their procedures** so that the rigor can be independently assessed **no matter what method is used**.

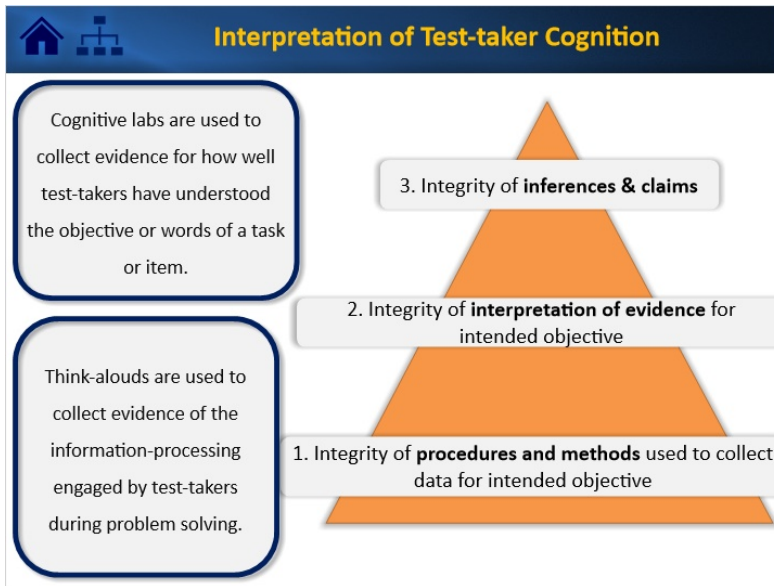
## ***2.15 Bookend: Test Development***



## ***2.16 Bookmark: Task and Method Alignment***



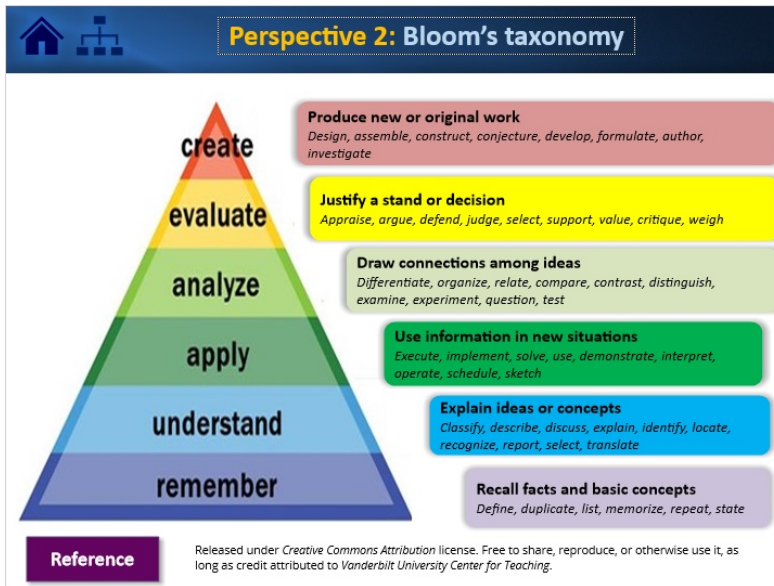
## 2.17 Interpretation of Test-taker Cognition



## 2.18 Research Goals

|  | Think-alouds  | Cognitive Labs  |
|--|---|---|
| <b>Test Development</b><br>Verbal reports are formative / exploratory to develop a cognitive model     |   |   |
| <b>Construct Validation</b><br>Verbal reports are summative / confirmatory to verify a cognitive model |   |   |

## 2.19 Bloom's Taxonomy



### Reference (Slide Layer)



**Reference**

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York and Toronto: Longmans, Green.

**Back**





## 2.20 Memory Demands



### Perspective 3: Memory Demands

- Although **human thinking** involves accessing **long-term memory** and **working memory**, depending on the nature of the task, memory systems **share the load differentially**.
- **Problem-solving tasks** often require application, analysis, evaluation and creation. These tasks **make high demands on working memory**.
- **Comprehension tasks** often require remembering and understanding. The tasks make **high demands on long-term memory**.

## 2.21 Method Choice



### Cognition and Task Differences

Problem-solving  
=  
Think-alouds

Task Comprehension  
=  
Cognitive Labs

View suggestions for how to collect data for these two different task types by clicking on the buttons.

## Problem solving (Slide Layer)

### Problem-solving = Think-alouds

If you want to collect data about how test-takers **solve problems**,  
then you are tapping their **working memory**

The interview method you use should include:

- ✓ Conditions for **observing** the generation of a solution
- ✓ Interview probes that **minimize leading or cueing or biasing** how test-takers select their problem solving strategies
- ✓ Focus on the **process** of problem solving as it happens

Back

## Task comprehension (Slide Layer)

### Task Comprehension = Cognitive Labs

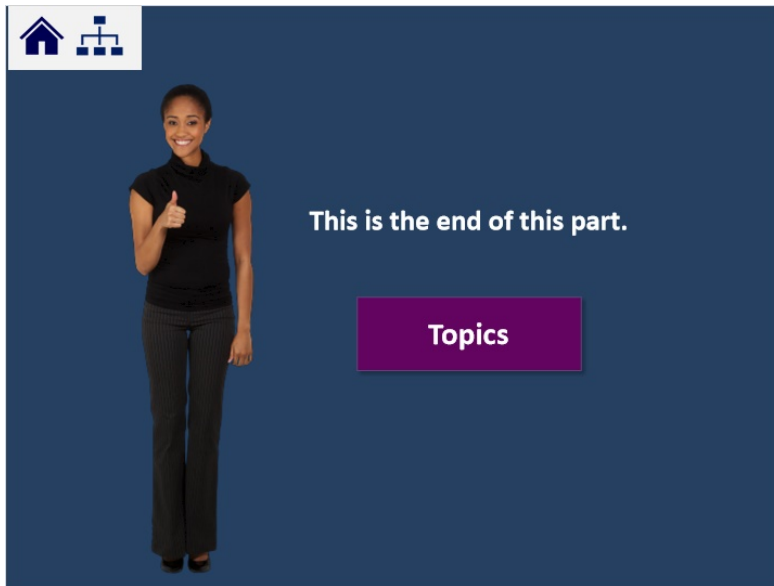
If you want to collect data about how test-takers **comprehend tasks**,  
then you are tapping **long-term memory** because  
you are not asking them to solve anything

The interview method you use should include:

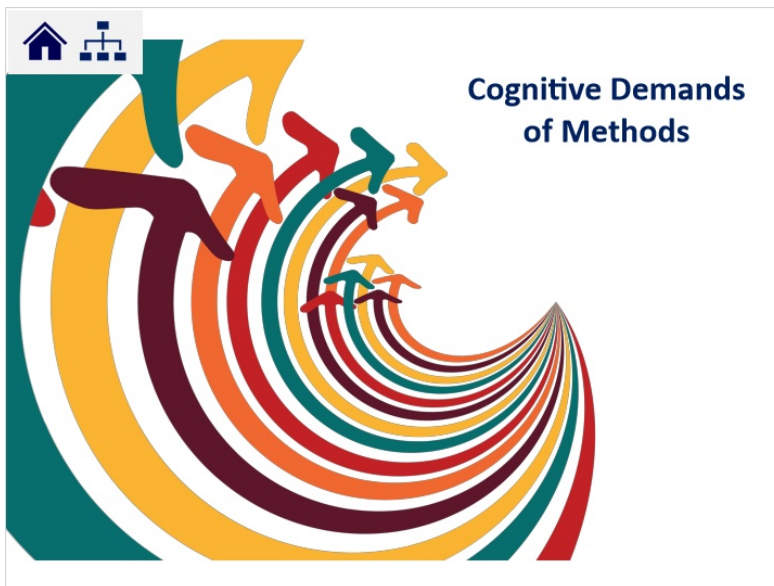
- ✓ Conditions for **observing** the generation of meaning
- ✓ Interview probes that **maximize how test-takers elaborate** on their understanding
- ✓ Focusing on the process and **consolidation** of understanding

Back

## 2.22 Bookend: Section 1



## 2.23 Bookmark: Memory and Cognition



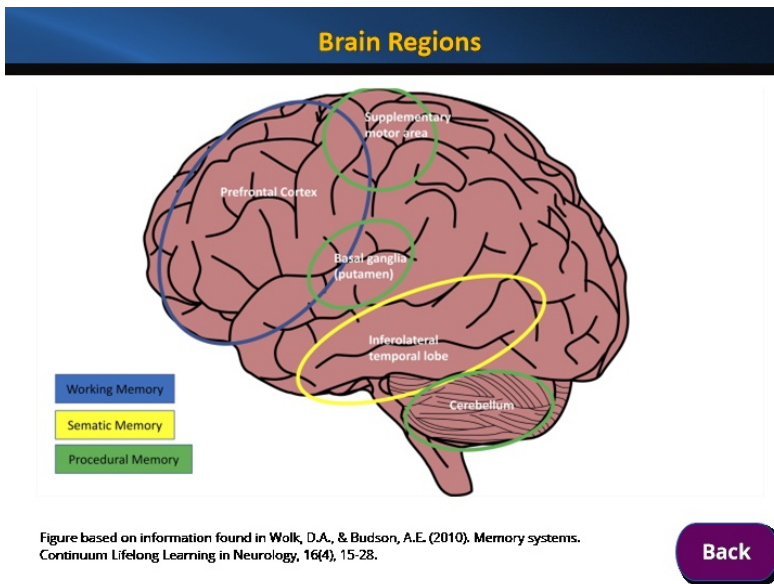
## 2.24 Memory System

| Memory Systems   |  |                                     |   |   |
|------------------|--|-------------------------------------|---|---|
| Memory Type      | Examples   | Aware to Report                     | Major Brain Structure/ Location   | Storage Length  |
| Episodic (LTM)   | Remembering the essay you wrote last week; the visit with your uncle and aunt  | Yes – controlled recall from LTM    | Medial temporal lobe, anterior thalamic nucleus, mamillary body, fornix, prefrontal cortex      | Minutes to years  |
| Semantic (LTM)   | Remembering the difference between smart and wise; why bears hibernate   | Yes – controlled recall from LTM    | Inferior lateral temporal lobes   | Minutes to years  |
| Procedural (LTM) | Remembering how to ride a bike; remembering how to log on to your old computer   | No - automated                      | Basal ganglia, cerebellum, supplementary motor areas  | Minutes to years  |
| Working          | Verbal based – repeating the code sent to you in order to log into an account<br><br>Spatial based – Observing the instructor work out a problem on the computer | Yes – controlled manipulation in WM | Prefrontal cortex, Broca and Wernike areas (for verbal), visual association areas (for spatial) | Seconds to minutes; unless information actively manipulated (e.g., repeating) |

Reference

Brain Regions

### Brain (Slide Layer)



## Reference (Slide Layer)

### Reference

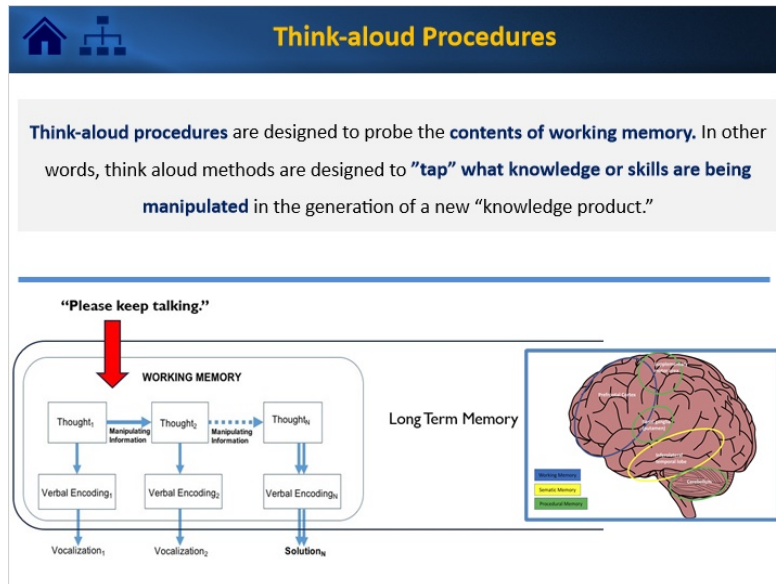


Table and figure based on:

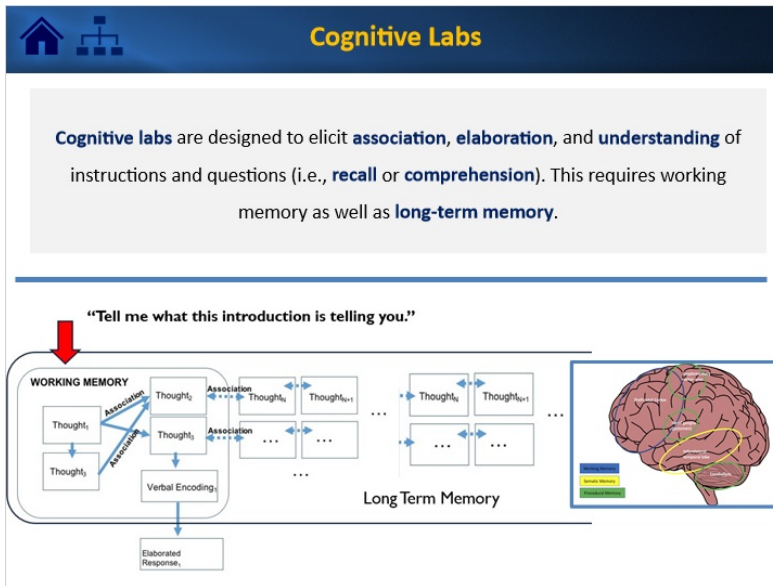
Wolk, D.A., & Budson, A.E. (2010).  
Memory systems. *Continuum  
Lifelong Learning in Neurology*,  
16(4), 15-28.

Back

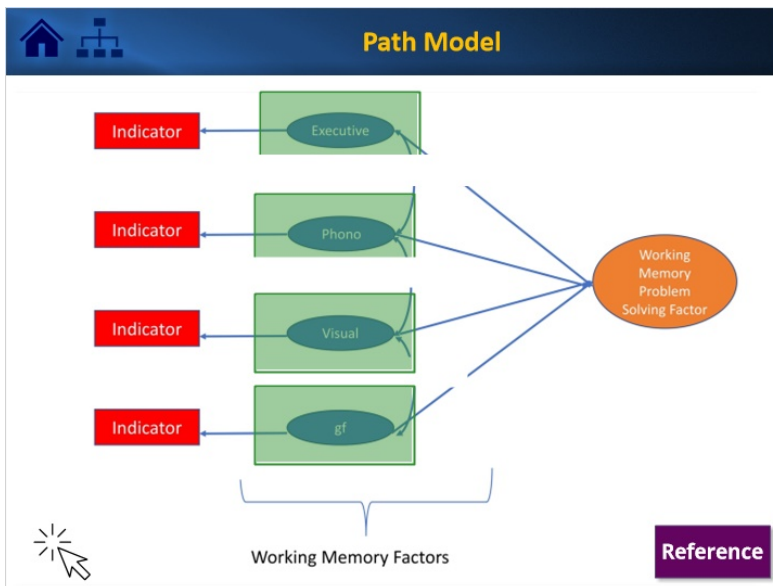
## 2.25 Think-alouds



## 2.26 Cognitive Labs



## 2.27 Path Model Example





## Reference (Slide Layer)

### Reference

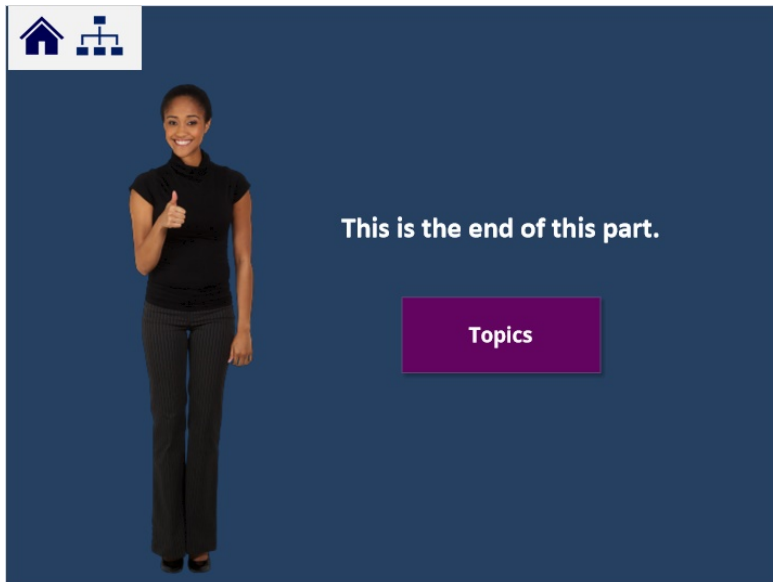


Figure on slide adapted / based on Figure 2 on p. 812 in:

Fung, W. & Swanson, H.L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition*, 45, 804-823.

Back

## 2.28 Bookmark: Memory and Cognition



### 3. Section 2: Collecting Verbal Reports

#### 3.1 Cover: Section 2



#### 3.2 Objectives: Section 3

**Learning Objectives**

1. Describe the different procedures for interview methods
2. Apply what is known about cognition to the selection of interview probes
3. Critically analyze the rigor of data for reflecting a type of cognition
4. Evaluate the credibility of claims made from verbal reports

### 3.3 Topic Selection





### 3.4 Bookmark: Method Differences



### 3.5 Overview

General Differences

| Think-alouds   | Cognitive Labs   |
|--|--|
| <ul style="list-style-type: none"> <li>• more specific timing of probes</li> <li>• minimal probing</li> <li>• measuring construction more than recall</li> </ul> | <ul style="list-style-type: none"> <li>• less specific timing of probes</li> <li>• elaborate probing</li> <li>• measuring recall more than construction</li> </ul> |

### 3.6 Research Goals

Research Goals

|  | Think-alouds | Cognitive Labs |
|--|--------------|----------------|
| <b>Test Development</b><br><br>Verbal reports are formative / exploratory to develop a cognitive model     |              |                |
| <b>Construct Validation</b><br><br>Verbal reports are summative / confirmatory to verify a cognitive model |              |                |

Click on each of the four cells (with checkmarks and circles) to learn more.

## 1 (Slide Layer)

### Think-alouds: Test Development

- Using **think-alouds for test / survey development** makes sense only if the objective is to explore the **construct** measured by the test / survey.
- The think aloud procedure is **too restrictive** in its probes.
- The main probe is **"Please keep talking"** so that the contents of working memory are **not biased** by elaborative probes.
- This is how **many testing programs** ought to be using think-alouds in the **initial stages of item development**.

Back

## 2 (Slide Layer)

### Cognitive Labs: Test Development

- Using **cognitive labs for test / survey development** makes more sense than for score claim validation
- The opportunity to **ask multiple questions** to find out how respondents are understanding tasks is precisely what is needed.
- This is how **Tourangeau (1984)** and **Willis (2015)** recommend to use cognitive labs.

Back

### 3 (Slide Layer)

#### Think-alouds: Construct Validation

- Using **think alouds for construct validation** makes sense if the objective is to identify key problem-solving response processes.
  - The procedure is structured to **minimize bias** in how interviewer and respondent interact.
  - The contents of **working memory** are being reported as this is where content is manipulated.
- This is **Ericsson and Simon (1993)** recommend the use of think-alouds.

Back

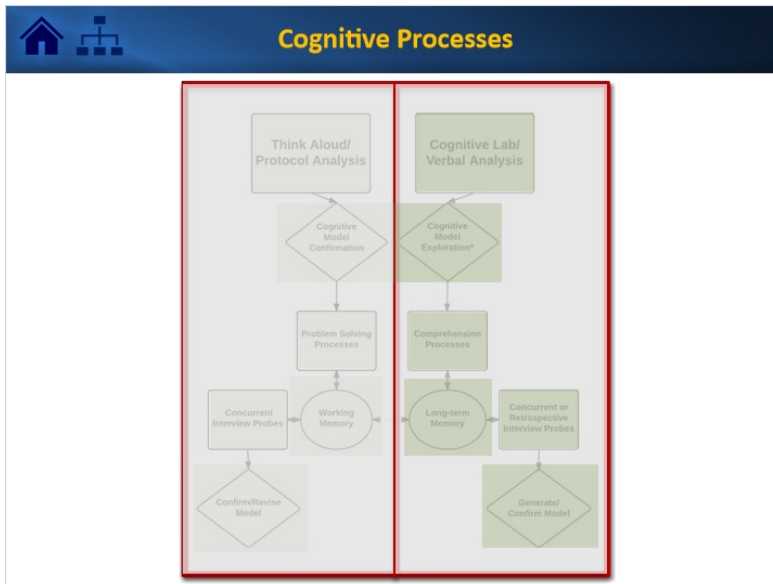
### 4 (Slide Layer)

#### Cognitive Labs: Construct Validation

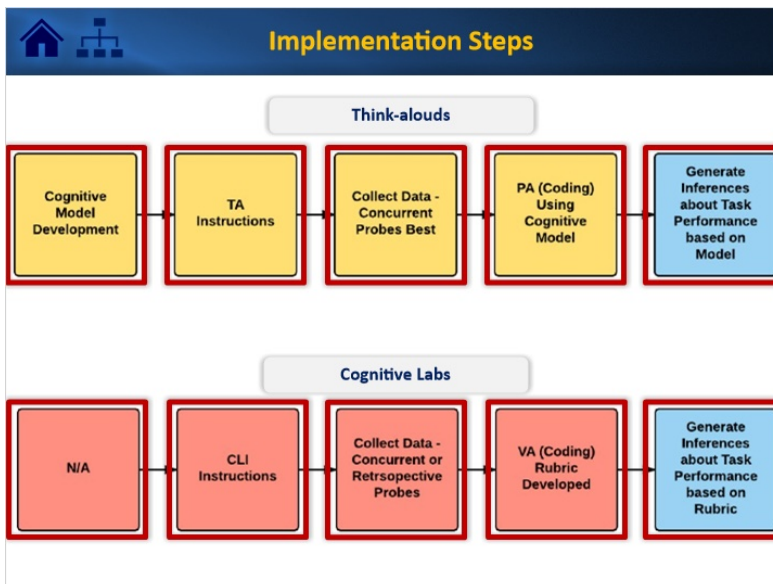
- Using **cognitive labs for construct validation** only makes sense if the objective is to map out conceptual or knowledge or comprehension differences between experts and novice test-takers.
- An inference cannot be made about how respondents solve tasks, only about **how their knowledge is structured** in relation to the task.
- This is how **Chi (1997)** and **Hmelo and Pfeffer (2004)** use cognitive labs to **map out conceptual differences**.

Back

### 3.7 Cognitive Processes





### 3.8 Implementation Steps






### 3.9 Summary I






Click on each row to learn more.



|              | Cog Labs                       | Think Alouds                   |
|--------------|--------------------------------|--------------------------------|
| Goal         | Comprehension                  | Problem solving                |
| Instruction  | Process of understanding       | Reporting of thinking process  |
| Timing       | More flexible timing           | Less flexible timing           |
| Length       | Depend on task and participant | Depend on task and participant |
| Participants | Smaller group                  | Larger group                   |
| Coding       | Grounded theory                | Cognitive model                |
| Agreement    | Required for thematic analysis | Required for coding analysis   |

#### Goal (Slide Layer)

Back

|      | Cog Labs   | Think-alouds  |
|------|--|---|
| Goal | Exploration of comprehension of pilot task or instrument;<br>level of task difficulty is not a concern | Testing of problem solving process in line with a pre-existing cognitive model of instrument;<br>level of task difficulty must be moderate for respondents to engage in problem solving |

Goal

Instruction

Timing



Length

Participants

Coding

Agreement

## Instruction (Slide Layer)



Back

|                    | Cog Labs  | Think-alouds   |
|--------------------|---|--|
| <b>Instruction</b> | Instructions emphasize process of understanding and respondent is made aware of the conversational quality of the interview | Instructions emphasize reporting of thinking process as it enters awareness and respondent is made aware of spartan probes used during the interview |

Goal

Instruction

Timing

Length

Participants

Coding

Agreement

## Timing (Slide Layer)



Back

|               | Cog Labs   | Think-alouds  |
|---------------|--|---|
| <b>Timing</b> | Verbal reports can be collected as the respondent is working on a task (concurrent) or after the respondent has finished (retrospective); response timing is not a concern | Verbal reports are best collected as the respondent is working on a task (concurrent); second best is immediately after the respondent has finished (retrospective) |

Goal

Instruction

Timing



Length

Participants

Coding

Agreement

## Length (Slide Layer)



Back

|               | Cog Labs  | Think-alouds  |
|---------------|---|---|
| <b>Length</b> | Interview length dependent on task, participant fatigue consideration | Interview length dependent on task, participant fatigue consideration |

Goal

Instruction

Timing



Length

Participants

Coding

Agreement

## Participants (Slide Layer)



Back

|                     | Cog Labs   | Think-alouds  |
|---------------------|--|---|
| <b>Participants</b> | Verbal report data are collected from small group of participants since inferences are descriptive and designed to provide formative assessment for tool improvement (goal part) | Verbal report data are collected from larger group of participants since inferences are inferential and designed to provide summative assessment of adequacy of cognitive model (goal part) |

Goal

Instruction

Timing


Length

Participants

Coding

Agreement

## Coding (Slide Layer)



Back

|               | Cog Labs  | Think-alouds   |
|---------------|---|--|
| <b>Coding</b> | Verbal report data coded using grounded theory, verbal analysis or other data-driven technique to generate themes | Verbal report data coded using cognitive model, protocol analysis or other theory-driven technique to generate codes |

Goal

Instruction

Timing


Length

Participants

Coding

Agreement

## Agreement (Slide Layer)



Back

|                  | Cog Labs   | Think-alouds   |
|------------------|--|--|
| <b>Agreement</b> | Inter-rater agreement is still required to generate credible conclusions about thematic analysis and conclusions | Inter-rater agreement mandatory to generate credible conclusions about coding analysis and conclusions |

Goal

Instruction

Timing



Length

Participants


Coding

Agreement

### 3.10 Summary II



Click on rows 2-4 to learn more.




|                      | Cog Labs  | Think-alouds                        |
|----------------------|---|-------------------------------------|
| <b>Model</b>         | Can be implemented without a pre-existing model | A pre-existing model is recommended |
| <b>Interviews</b>    | Easy to conduct                                 | Hard to conduct                     |
| <b>Data Analysis</b> | Exploratory                                     | Confirmatory                        |
| <b>Inferences</b>    | depend on task and participant                  | depend on task and participant      |

#### Interviews (Slide Layer)

Back

|                   | Cog Labs  | Think-alouds  |
|-------------------|---|---|
| <b>Interviews</b> | Interviews are easier to conduct because they are:<br><br>(a) conversational and timing of probes is less problematic.<br>(b) tasks can be of varying difficulty. | Interviews are sometimes awkward to conduct because:<br><br>(a) they are not conversational and timing of probes is consequential.<br>(b) tasks must be of moderate difficulty for respondents to |

InterviewsDataInferences




## Data Analysis (Slide Layer)

[Back](#)

|                      | Cog Labs   | Think-alouds  |
|----------------------|--|---|
| <b>Data Analysis</b> | <p>Data analysis is purely descriptive given small sample size.</p> <p>Data analysis can be time-consuming because coding rubric of themes is defined from the breadth</p> | <p>Data analysis is inferential given that the assessment of a model is driving the data collection and therefore larger, more representative samples have to be recruited.</p> <p>Data analysis is typically not time-</p> |

[Interviews](#)[Data](#)[Inferences](#)



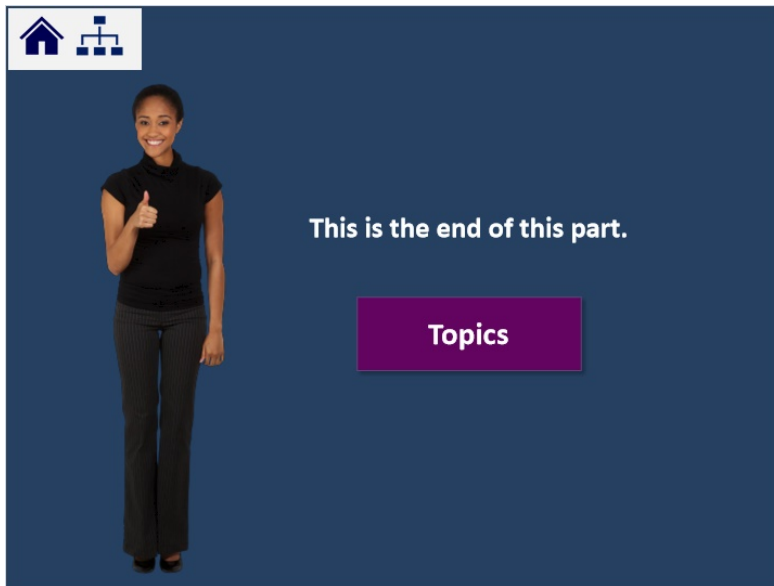
## Inferences (Slide Layer)

[Back](#)

|                   | Cog Labs   | Think-alouds  |
|-------------------|--|---|
| <b>Inferences</b> | <p>Inferences are confined to claims about response processes associated with knowledge, understanding and/or comprehension.</p> | <p>Inferences are narrow and confined to claims about response processes associated with problem solving given the conditions under which the respondent solved the task.</p> |

[Interviews](#)[Data](#)[Inferences](#)

### ***3.11 Bookend: Method Differences***



### ***3.12 Bookmark: Traditional Methods***







### 3.13 Think-alouds I



### 3.14 Think-alouds II




### 3.15 Cognitive Labs



Reflection Question



What might this process look like for **cognitive labs**?  
Write down what the **interviewer might say**.

When you are ready, click on the **EXAMPLE** button below  
to hear an audio explanation.



Example


### 3.16 Interview Concerns I





Interview Concerns

Interviews are used to collect **self-reported information** from respondents and are therefore **subject to well-known errors** that can corrupt the quality of the data gathered:

- ✓ social desirability
- ✓ forgetting/demotivation
- ✓ interviewer bias
- ✓ interaction process



### 3.17 Interview Concerns II





#### Interviewer-respondent Bias

- Because the interview involves **two human beings in a one-to-one interaction**, the interviewer and respondent each brings **potential bias** to the interaction.
- The bias can be present even when working from an **interview script** since communication involves **non-verbal behavior**.

**Examples:**

- The interviewer may use **verbal or non-verbal cues** that influence the respondent's performance, comfort and/or motivation (Leighton, 2013)
- If the respondent holds **less constructive behavioral attitudes** towards mistakes and/or **feels uncomfortable** during the interview, this has been found to **depress performance** (Lutsyk & Leighton, 2018)

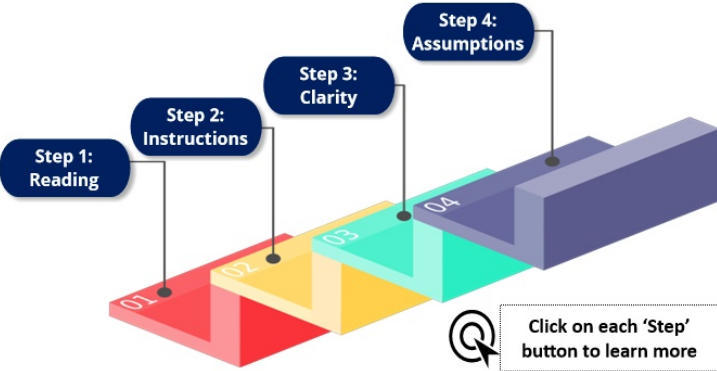
### 3.18 Probe Quality



#### Quality of Probes

Willis (2015) provides examples of interview probes for survey items and associated quality-control criteria for the probes.

**Reference**



Click on each 'Step' button to learn more

## Step 1 (Slide Layer)

**Step 1: Reading** [Back](#)

| Determine whether it is difficult for the interviewers to read the question in the same way to all respondents                 |  |
|--|--|
| <b>1a. WHAT TO READ:</b><br>Interviewer may have difficulty determining what parts of the question should be read.             | <b>PROBE:</b><br>None. Interviewer should attend to this potential difficulty and note it. |
| <b>1b. MISSING INFORMATION:</b> Information the interviewer needs to administer the question is not contained in the question. | <b>PROBE:</b><br>Interviewer should attend to this potential difficulty and note it.       |
| <b>1c. HOW TO READ:</b><br>Question is not fully scripted and therefore difficult to read.                                     | <b>PROBE:</b><br>Interviewer should attend to this potential difficulty and note it.       |

[Step 1: Reading](#) [Step 2: Instructions](#) [Step 3: Clarity](#) [Step 4: Assumptions](#)

## Step 2 (Slide Layer)

**Step 2: Instructions** [Back](#)

| Look for problems with any introductions, instructions, or explanations from the respondent's point of view.     |  |
|--|--|
| <b>2a. CONFLICTING OR INACCURATE INSTRUCTIONS:</b><br>Introductions or explanations are problematic / confusing. | <b>PROBE:</b><br>(At the end of a long intro, but before the question itself): "Before I get to the actual question, tell me what this introduction is telling you." |
| <b>2b. COMPLICATED INSTRUCTIONS:</b><br>Introductions or explanations are problematic / confusing.               | <b>PROBE:</b><br>Same as above.  |

[Step 1: Reading](#) [Step 2: Instructions](#) [Step 3: Clarity](#) [Step 4: Assumptions](#)

## Step 3 (Slide Layer)

**Step 3: Clarity**[Back](#)

Identify problems related to conveying the *intent / meaning* of the question to the respondent.

|   |  |
|---|--|
| <b>3a. WORDING:</b><br>Question is lengthy, awkward, ungrammatical, or contains complicated syntax.           | <b>PROBE:</b><br>"Can you tell me in your own words what that question was asking?"  |
| <b>3b. TECHNICAL TERM(S):</b><br>Are undefined, unclear, or complex.  | <b>PROBE:</b><br>"What does the word (term) mean to you, as it's used in this question?"   |
| <b>3c. VAGUE:</b><br>There are multiple ways to interpret the question or decline what to include or exclude. | <b>PROBE:</b><br>"Tell me what you were thinking when I asked about (topic)."  |
| <b>3d. REFERENCE PERIODS:</b><br>Are missing, not well specified, or in conflict.                             | <b>PROBES:</b><br>(1) "Can you remember what time period the question was asking about?"<br>(2) "You said (answer). What time period does that cover?" |

Step 1:  
Reading

Step 2:  
Instructions

Step 3:  
Clarity

Step 4:  
Assumptions

## Step 4 (Slide Layer)

**Step 4: Assumptions**[Back](#)

Determine if there are problems with assumptions made or the underlying logic.

|  |   |
|--|---|
| <b>4a. INAPPROPRIATE ASSUMPTIONS:</b><br>Are made about the respondent or about his or her living situation. | <b>PROBES:</b><br>(1) "How well does that question apply to you?"<br>(2) "Can you tell me more about that?" |
| <b>4b. ASSUMES CONSTANT BEHAVIOR:</b><br>Or experiences that vary.   | <b>PROBE:</b><br>"Would you say that mostly stays the same, or does it vary or depend?"                     |
| <b>4c. DOUBLE-BARRELED:</b><br>Contains more than one implicit question.                                     | <b>PROBE:</b><br>"Tell me more about your opinions on that."  |

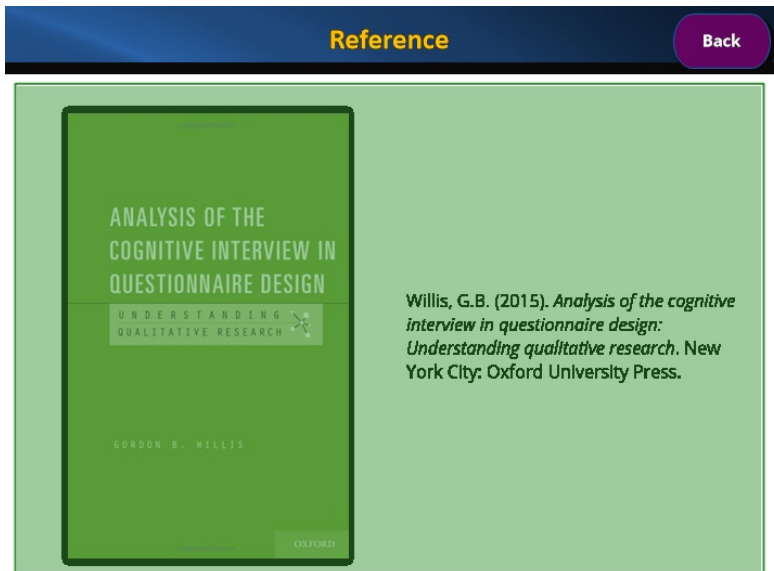
Step 1:  
Reading

Step 2:  
Instructions

Step 3:  
Clarity

Step 4:  
Assumptions

## Reference (Slide Layer)

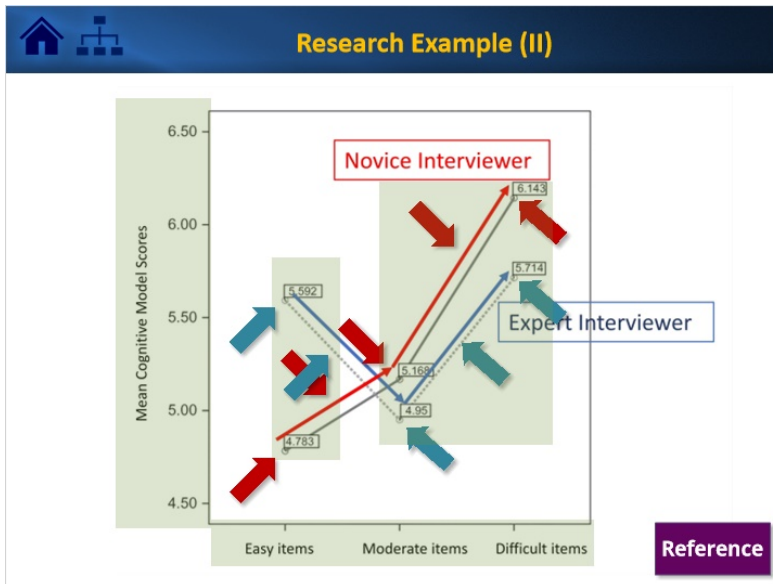


### 3.19 Example I

The slide has a dark blue header with a home icon and a tree diagram icon, and the title "Research Example (I)" in yellow. The main content area is white. It starts with a paragraph: "Leighton (2013) found that even a **few introductory sentences** at the beginning of an interview could **alter respondents' thinking process**:". Below this are three bullet points, each in a light blue box with a dark blue border:

- Interviewers who introduced themselves **as experts in the area of the study** or **said nothing at all about their expertise** had the effect of lowering respondents' scores on a think-aloud task involving math items.
- Respondents who thought they were being interviewed by a non-expert **showed a different pattern of retrospective reports** for their item solutions than respondents who thought they were reporting to an expert
- **Computer-based interview tools** offer a promising approach to minimizing bias by controlling the interviewer-respondent interactions.

### 3.20 Example II



### Reference (Slide Layer)

### Reference

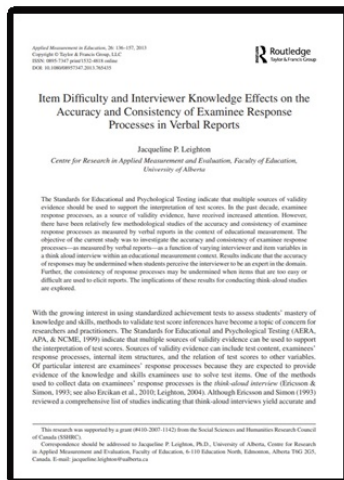


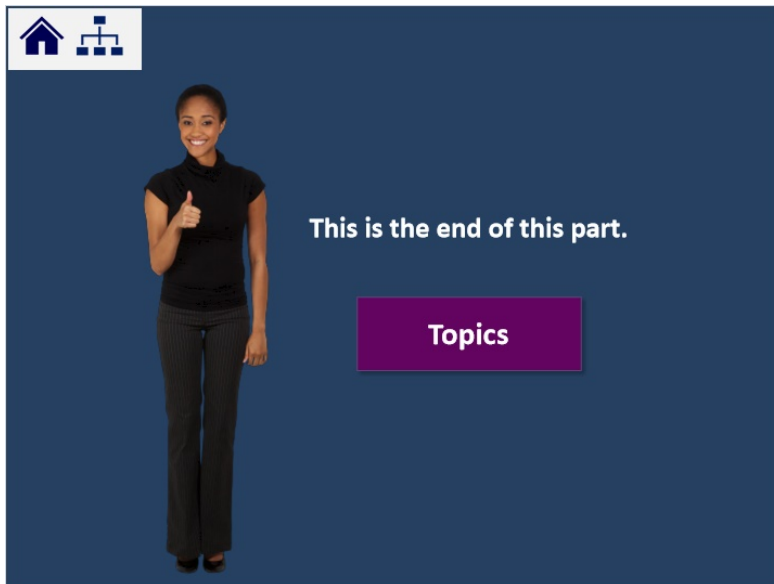
Figure adapted/based on:

Leighton, J.P. (2013). Item difficulty and interviewer knowledge effects on the accuracy and consistency of examinee response processes in verbal reports. *Applied Measurement in Education*, 26, 136-157.

Back



### 3.21 Bookend: Traditional Methods





## 4. Section 3: Analyzing Verbal Reports


### 4.1 Cover: Section 3



## 4.2 Objectives: Section 4

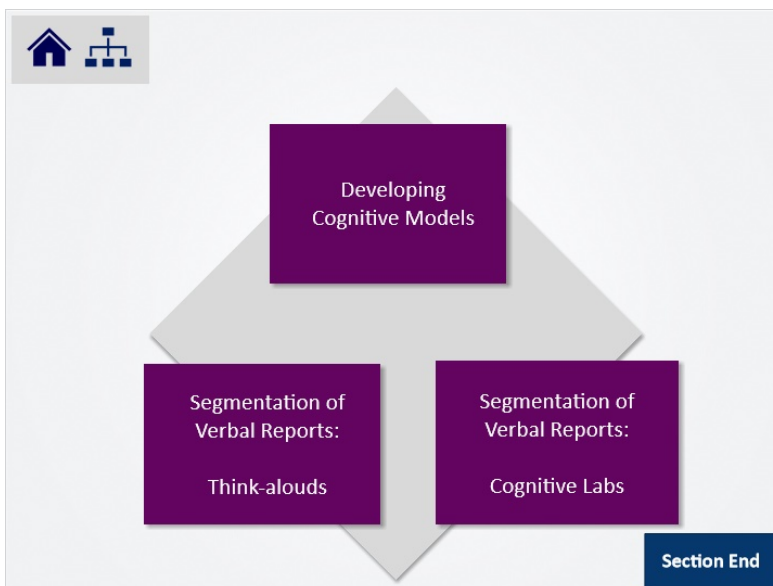


### Learning Objectives



1. Describe model development for think aloud interview studies
2. Apply distinct segmentation and coding techniques
3. Critically analyze examples of think aloud and cognitive lab methods
4. Evaluate credibility of data contingent on inter-rater agreement



## 4.3 Topic Selection



## 4.4 Bookmark: Model Development



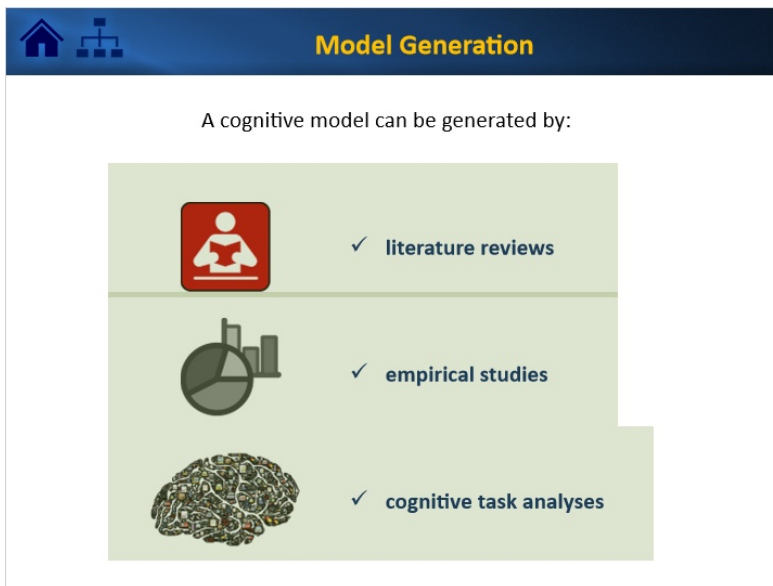
## 4.5 Cognitive Models vs Exploratory Model Generation I

**Rationale**

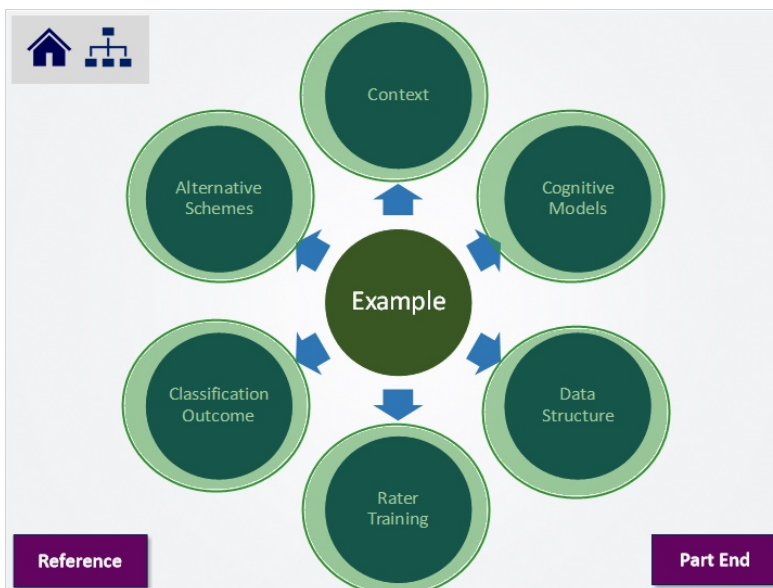
According to Ericsson and Simon (1993), the reason to collect **think-aloud interviews** involves the **testing of cognitive models**, which involves:

- A **theoretical hypothesis** of the sequence of processes expected to be observed in producing the correct response (and/or incorrect depending on what sequence of processes are outlined in the model)
- A **coding artifact**, where aspects of the model(s) reflect potential codes, which are then summarized in a coding book to guide analysis of verbal reports.

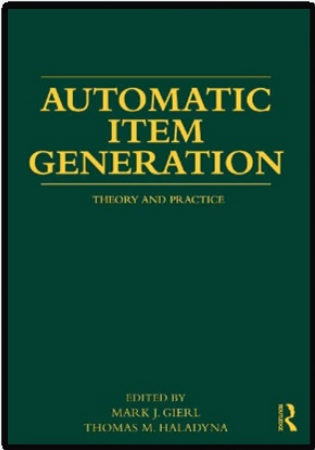
## 4.6 Model Generation II



## 4.7 Example Part Selection





## Reference (Slide Layer)



Leighton, J.P. Learning sciences, cognitive models, and automatic item generation. (2013). In M.J. Gierl & T. Haladyna (Eds.), *Automatic item generation: Theory and practice* (pp. 121-135). Routledge.

Back

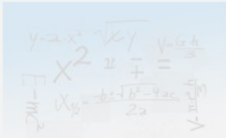
## 4.8 Example 1




Example: Study Context

Back to Graphic

Leighton (2013) asked content experts in Grade 12 mathematics to illustrate the relevant steps required to respond to 15 pre-calculus mathematics items.

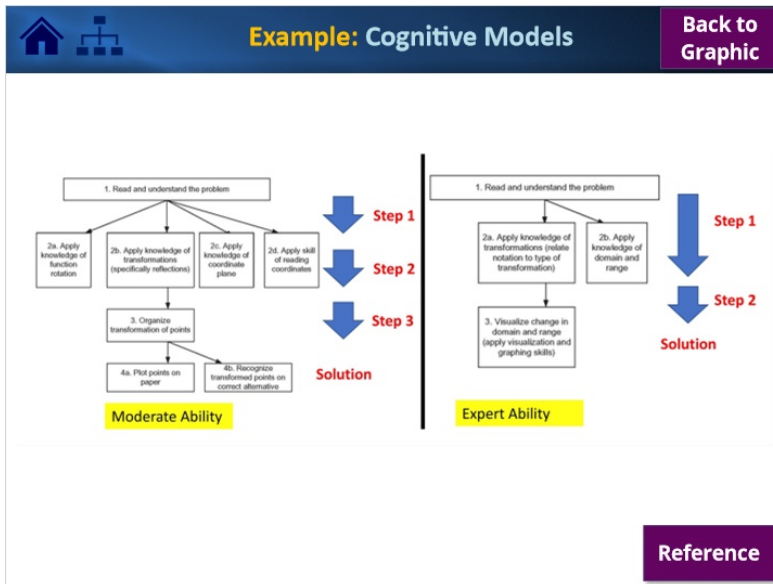




The content experts were asked to generate two cognitive models

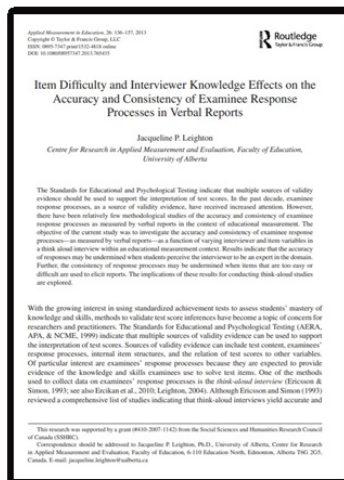
1. expert model
2. moderate-ability model

## 4.9 Example II



## Reference (Slide Layer)

## Reference





**Figure adapted / based on Information In:**

Leighton, J.P. (2013). Item difficulty and interviewer knowledge effects on the accuracy and consistency of examinee response processes in verbal reports. *Applied Measurement in Education*, 26, 136-157.

Back

#### 4.10 Example III




Example: Data Structure



Back to Graphic

A total of **60 cognitive models** (15 items x 2 content experts x 2 models) were developed based on task analyses to guide the coding of **1,065 full verbal reports** (71 students x 15 items).

Each of the 1,065 full verbal reports included a **concurrent report** and a **retrospective report** for a total of 2,130 specific reports.




#### 4.11 Example V




Example: Rating Process

Back to Graphic



**Rater 1:** Independent and Blind to Study



| Phase 1 (Train)          | Phase 2 (Assess) | Phase 3 (Final Codes)    |
|--------------------------|------------------|--------------------------|
| 45 concurrent reports    | Cohen's kappa    | 1,020 concurrent reports |
| 45 retrospective reports | >.80             | 1,020 retrospect reports |



**Rater 2:** Independent and Blind to Study



#### 4.12 Example IV



Example: Classification Outcome

Back to Graphic



2 independent raters, blind to the nature of the study,  
holistically categorized the 2,130 concurrent and retrospective verbal reports  
into 1 of 5 categories:

|     |                         |                    |
|-----|-------------------------|--------------------|
| (1) | Model A                 | (high ability)     |
| (2) | Model B                 | (high ability)     |
| (3) | Model C                 | (moderate ability) |
| (4) | Model D                 | (moderate ability) |
| (5) | No match with any model |                    |

#### 4.13 Bookmark: Segmentation for Think-alouds




## 4.14 Segmentation of Verbal Report



### Segmentation of Verbal Reports

In **think aloud interviews**, the respondent is solving a task and reporting thoughts about problem solving. The verbal reports are therefore **continuous narratives with intermittent probes** to “Keep talking.”  
(Lutsyk & Leighton, 2018)



Sample Item & Verbal Report

### Example (Slide Layer)

Segmentation of Verbal Report

Back

Use the following information to answer the next question.

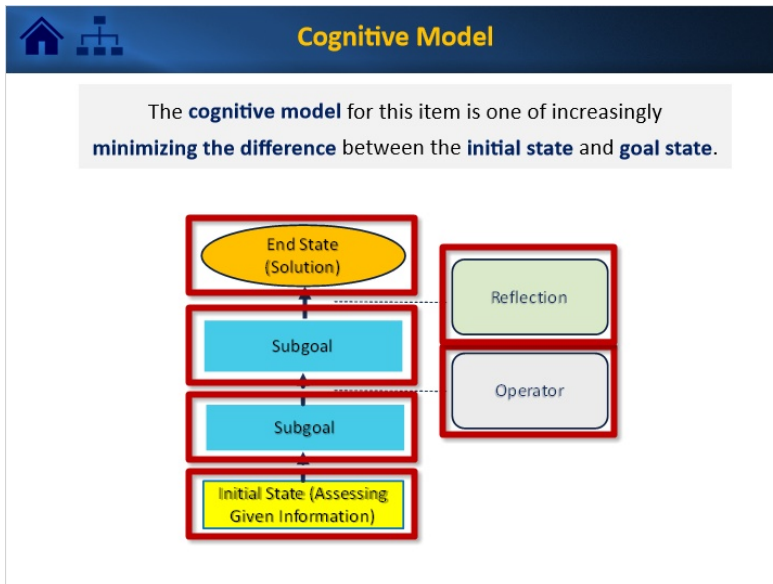
A candy dish contains 2 blue candies, 1 red candy, 3 white candies, and 2 green candies. Two candies will be randomly selected from the dish, one after the other, without replacement.

14. The probability of selecting a blue candy and then a candy that is not blue is

A.  $\frac{3}{14}$   
B.  $\frac{3}{16}$   
C.  $\frac{1}{16}$   
D.  $\frac{1}{28}$

6 out of 7 equals.... I mean,... therefore 7 equal 25%. And then divide 25 by 7 is 6 or 5, it's like 40, 12, 16, 29, 24, no, ??, umm.. 36 ??, 12, 18, 21, 24. It's 8, it's like 4 point something which leads me to the closest answer would be B, umm 3 out of 16.

## 4.15 Segmentation of Verbal Report



## 4.16 Example

**Example Coding**

|   |   |
|---|---|
| <i>I have zero math background.</i>   | Self-assessment about math knowledge<br><b>[Metacognition]</b>                    |
| <i>So, I have to find out how many candies there are total.</i>   | Judgment about part of problem to be solved. <b>[End State/ Goal State]</b>       |
| <i>So there are 2 blue candies, 1 red, 3 white, and 2 green. So, 2 B, 1 R, 3 W, 2 G. In total there are 2, 4, 5-6-7 in total 8 candies.</i> | Processing/ identification of given information and adding <b>[Initial State]</b> |

|   |                           |
|---|---------------------------|
| <b>Yellow</b> = No information manipulation | <b>Green</b> = Reflection |
| <b>Orange</b> = Self-assessment             | <b>Gray</b> = Operator    |
| <b>Blue</b> = Information processing        |                           |

## 4.17 Inter-rater Agreement and Credibility of Inferences



### Inter-rater Agreement

- Having **at least 2 raters** trained to **independently review and code** the verbal reports is important to establish **credibility of interpretation**
- This requires one to develop a **code book**, then **train raters** on 10-15% of the verbal reports, and to **iterate / revise** until the materials are **robust** before coding of the **remaining responses** is done
- Cohen's kappa variants** can be used as indices of **inter-rater agreement** to **monitor raters** and the associated **implementation of the procedure**



Formula

Thresholds

### formula (Slide Layer)

#### Formula

$$K = \frac{P_{\text{agree}} - P_{\text{chance}}}{1 - P_{\text{chance}}}$$

$P_{\text{agree}}$  = Proportion of trials in which judges agree

$P_{\text{chance}}$  = Proportion of trials in which agreement would be expected due to chance

Back

## Thresholds (Slide Layer)

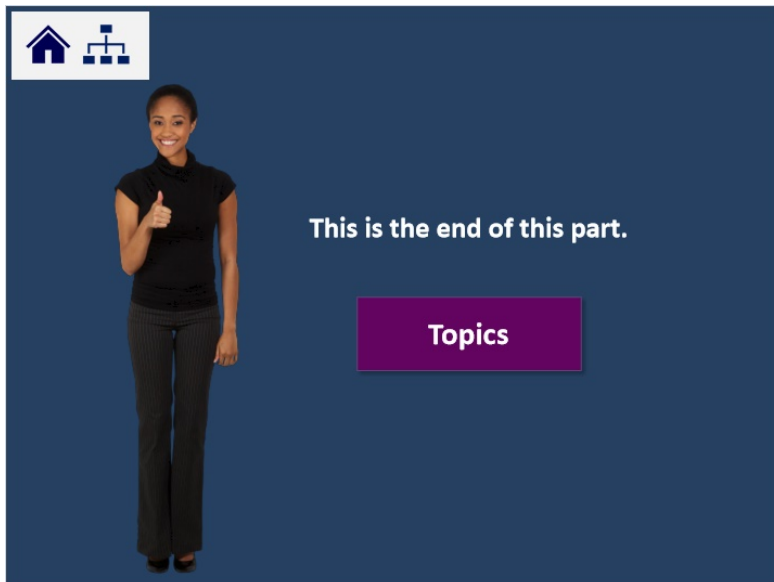
### Interpretation of Kappa Values

Landis and Koch (1970) suggest the following:



- Values less than or equal to 0 suggest **less than chance** agreement
- Values .01 to .20 suggest **slight** agreement
- Values 0.21 to 0.40 suggest **fair** agreement
- Values 0.41 to 0.60 suggest **moderate** agreement
- Values 0.61 to 0.80 suggest **substantial** agreement
- Values 0.81 to 0.99 suggest **almost perfect** agreement

Back

## 4.18 Bookend: Segmentation for Think-alouds




## 4.19 Coding Schemes



Coding Schemes

Wang and Leighton (2011) discuss  
four additional coding schemes ranging from **strict to lenient**:




Coding Scheme  
1

Coding Scheme  
2

Coding Scheme  
3

Coding Scheme  
4

Click on each button to view the scheme 

### Scheme 1 (Slide Layer)

Coding Scheme 1

Back

If, for a given item and a given student, the concurrent report and retrospective report **show identical cognitive model classification**, then a value of **1** is assigned for consistency, otherwise a **0**.

This coding was used to determine whether students' concurrent and retrospective reports were **completely identical in the cognitive model** reflected in their thinking processes.

Coding Scheme  
1

Coding Scheme  
2

Coding Scheme  
3

Coding Scheme  
4

## Scheme 2 (Slide Layer)

**Coding Scheme 2** [Back](#)

If, for a given item and a given student, the concurrent report and retrospective report show **a similar ability-level of model**, then a value of **1** was assigned for consistency, otherwise a **0**.

This coding was used to determine whether students' concurrent and retrospective reports **matched in terms of ability**.

[Coding Scheme 1](#) [Coding Scheme 2](#) [Coding Scheme 3](#) [Coding Scheme 4](#)

## Scheme 3 (Slide Layer)

**Coding Scheme 3** [Back](#)

If for a given item and a given student, the concurrent and retrospective report are categorized into one of the two models designed by **a single test developer**, then a value of **1** was assigned, otherwise a **0**.

This coding was used to determine whether students' concurrent and retrospective reports **matched in terms of the knowledge and skills** the test developers considered relevant for success on the item.

[Coding Scheme 1](#) [Coding Scheme 2](#) [Coding Scheme 3](#) [Coding Scheme 4](#)



## Scheme 4 (Slide Layer)

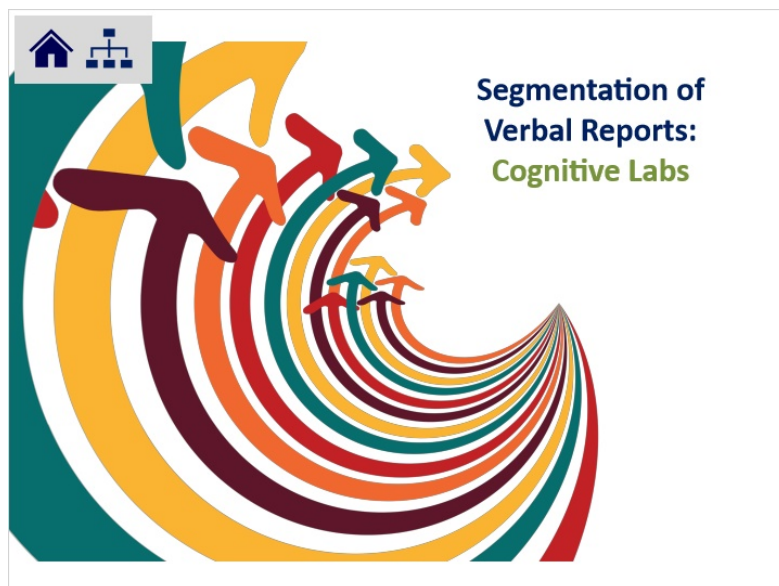
**Coding Scheme 4** [Back](#)

If for a given item and a given student, **both** concurrent and retrospective reports are categorized as showing any kind of model, then a value of **2** is assigned. If **either** the concurrent or retrospective are categorized into any kind of cognitive model, then a **1** is assigned. If **neither** the concurrent nor the retrospective are categorized as showing any cognitive models, then a **0** is assigned.



This coding was used to evaluate the degree to which students' concurrent and retrospective reports reflected **full use** of cognitive models or even **partial use** of models.

[Coding Scheme 1](#) [Coding Scheme 2](#) [Coding Scheme 3](#) [Coding Scheme 4](#)

### 4.20 Bookmark: Segementation for Cog Labs





## 4.21 Example





### Segmentation of Verbal Report

In **cognitive labs**, the respondent is relaying how he or she **understands or perceives** a question or item. The verbal reports are therefore **continuous conversations** with the interviewer as the respondent answers the interviewer's questions and **elaborates on probes** such as "Tell me more about what you're exactly thinking about."



## Example Item and Interview Questions (Slide Layer)



Use the following information to answer the next question.

A candy dish contains 2 blue candies, 1 red candy, 3 white candies, and 2 green candies. Two candies will be randomly selected from the dish, one after the other, without replacement.

14. The probability of selecting a blue candy and then a candy that is not blue is

A.  $\frac{3}{14}$

B.  $\frac{3}{16}$

C.  $\frac{1}{16}$



D.  $\frac{1}{28}$

3. Did you feel any emotions as you were solving the task?


4. Is there anything you would change about the task?

5. How difficult or easy was it to solve the task while voicing your thoughts?

## 4.22 Illustration I




### Illustration (I)




Now that you are finished with the task, I will ask you the five questions....

Can you tell me the steps or any strategies you used to solve the task?


Umm, [laughing, I don't know what the actual names of these strategies are, umm ... I, I used, I mean, I mean statistics, probability. What I should have done is to use multiple choice testing strategies and figure out which one of these are ridiculous. I don't think it's 1 out of 28. And so on and so forth.







Did you find the task difficult or easy or so-so?


Difficult, because I don't do math.



## 4.23 Illustration II





### Illustration (II)



Did you feel any emotions as you were solving the task?


I mean moderate frustration.






Is there anything you would change about the task?


If I was to give it to a student? The task in all seems fine.





How difficult or easy was it to solve the task while voicing your thoughts?

Maybe easier than not voicing my thoughts.  
It's different. I don't know, it's equal to compare.



## 4.24 Coding Complexity

Coding Complexity

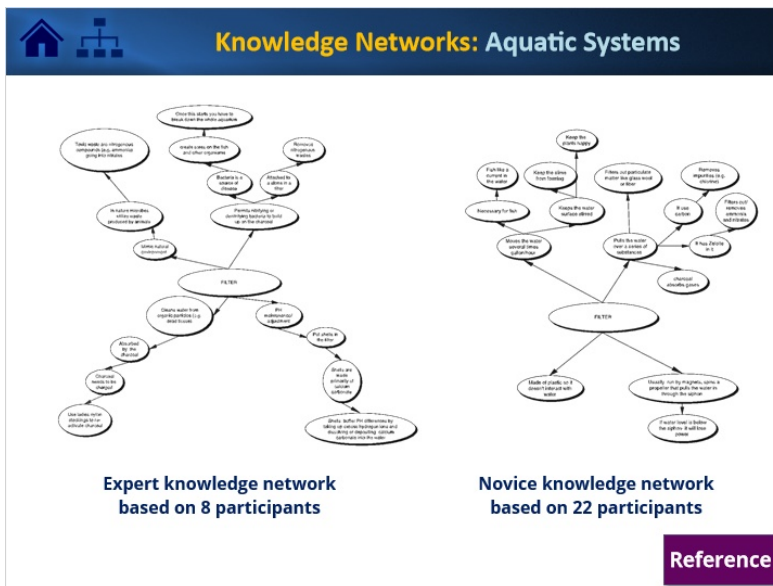
**Coding is guided by the structure of the questions:**

If interviewer questions provide **options / choices** for respondents, the analysis is a simple description of **how many respondents** chose the different options.

When the interview script involves **questions plus elaborative probes**, then using the interview script is a **small part of coding**. In this case, **themes and codes** are outlined along with exemplars.

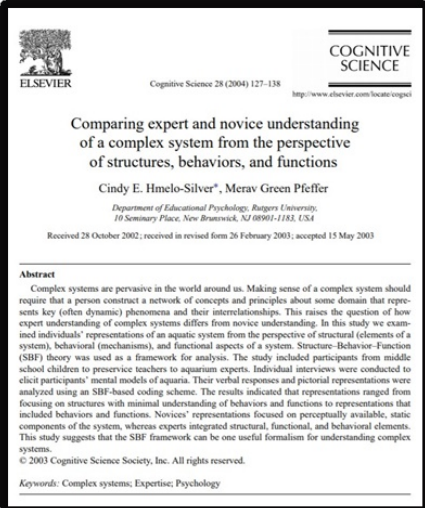
For either coding complexity, a **formal rubric** is developed to apply codes

## 4.25 Knowledge Networks



## Reference (Slide Layer)

### Reference





**Original figures 1 and 2  
reprinted with permission from  
Wiley from:**

Hmelo-Silver, C., & Pfeffer, M. G.  
(2004). Comparing expert and  
novice understanding of a  
complex system from the  
perspective of structures,  
behaviors, and functions.  
*Cognitive Science*, 28(1), 127-138.

**Back**

## 4.26 Code Generation

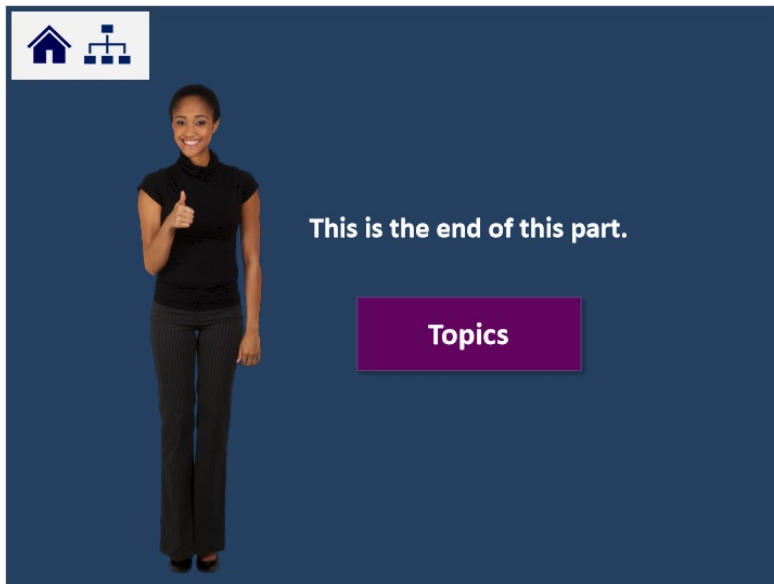


### Code Generation

**Code generation by raters is an especially time-consuming process because:**

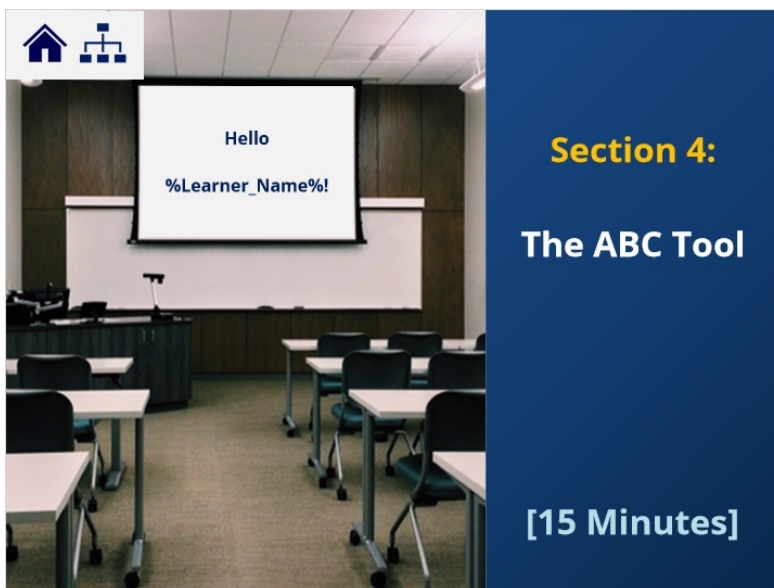
- Raters need to be trained to review reports to develop a list of themes and codes
- Ratings have to be assessed for comparable identification of themes and codes
- Interrater agreement statistics have to be calculated
- Changes may have to be made to the themes and codes
- More coding will need to be conducted if such changes are made
- Additional interrater agreement statistics will need to be re-calculated
- Only upon satisfactory completion, reports are coded and represented visually

#### 4.27 Bookend: Segmentation for Cog Labs





### 5. Section 4: ABC Tool

#### 5.1 Cover: Section 4



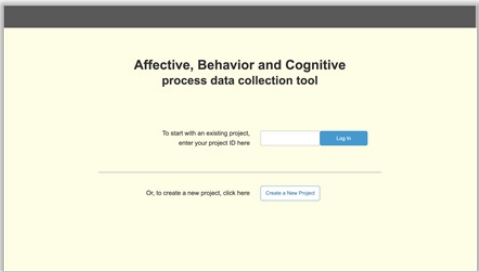
## 5.2 Overview



### Introduction

**Think-aloud interviews** and **cognitive labs** provide valuable information about **response processes** and **learner experiences / perceptions**

**Difficult to implement at scale** due to one-to-one match of interviewer and participants – **ABC tool** can help with scaling up





To get **access to the ABC tool** contact Blair Lehman at ETS ([blehman@ets.org](mailto:blehman@ets.org))


## 5.3 Team





### Design Team

**Tom Florek**

**Enruo Guo**

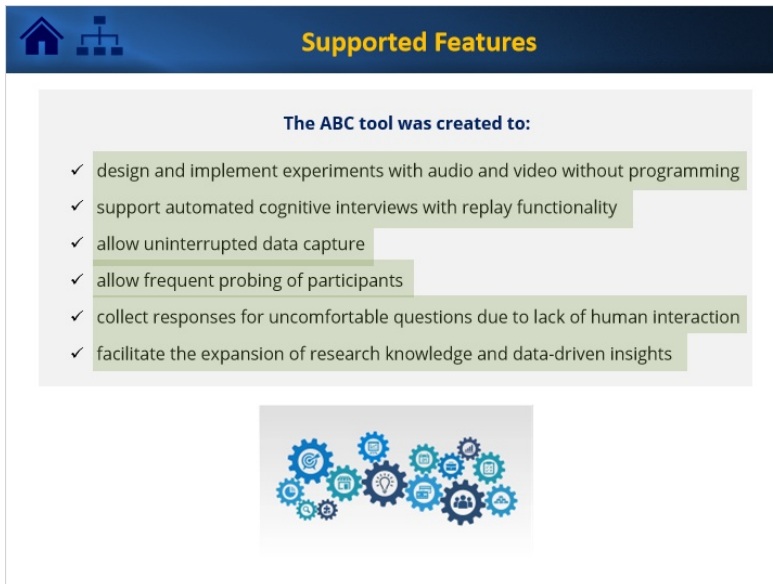
**Blair Lehman**

**Pavan Pillarisetti**

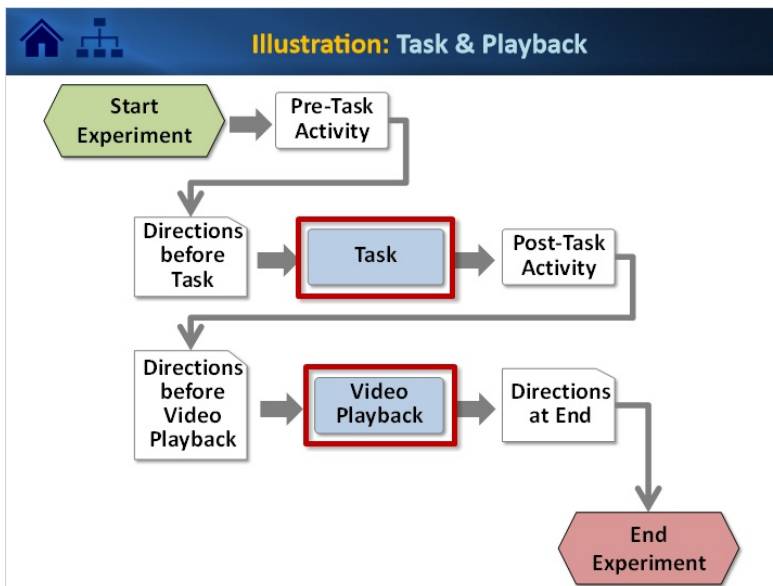
**Deb Pisacreta**



## 5.4 Supported Features



## 5.5 Experimental Design (I)



## 5.6 Playback Interface

The screenshot shows the 'Playback Interface' with a dark blue header containing a home icon and a tree diagram. The main content area is divided into three sections. The top-left section is a video player titled 'Filename: microgames\_fall2017\_MuVsn\_TestT02\_1.mpg' showing a cartoon titled 'Awkward Annie' with a score of 15 points and Level 2. The video shows two characters, a boy and a girl, with speech bubbles. The top-right section is a video of a person wearing a headset. The bottom section is a feedback form titled 'How were you feeling at this point in the interaction?' with a grid of 16 emotion options: Angry, Curious, Frustrated, Sad, Anxious, Delighted, Happy, Shame, Bored, Disgusted, Hopeful, Surprise, Confused, Engaged, Hopeless, Neutral, and Contempt. A 'continue' button is at the bottom right.

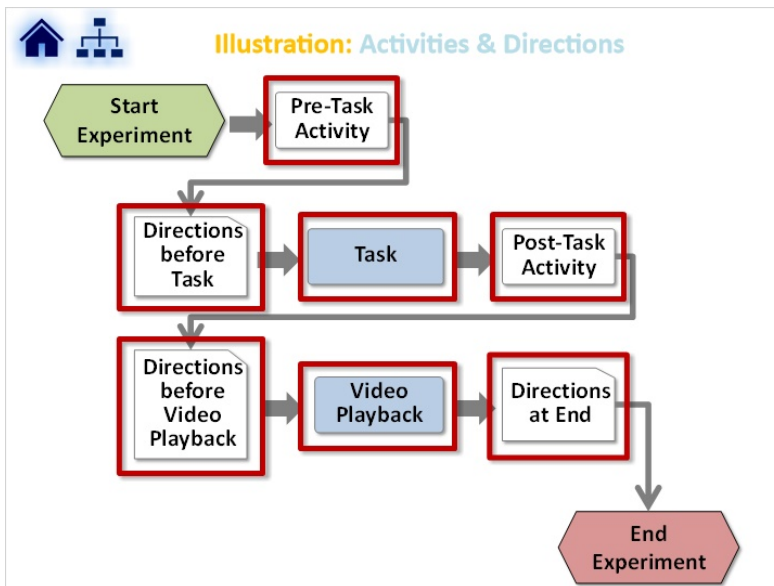
## 5.7 Question Functions

The screenshot shows the 'Question Functions' slide with a dark blue header containing a home icon and a tree diagram. The main content area is a light gray box with the title 'The following question features are included:' and a list of features:

- ✓ automatically pose questions at specific:
  - time intervals (e.g., every 30 seconds)
  - time locations (e.g., at 1:30 minutes or 5 minutes)
  - events (e.g., everyone responds to question 2)
- ✓ event insertion requires access to log files
- ✓ up to four questions can be inserted at each stop point
- ✓ questions can take multiple formats (e.g., MC, CR, grid)
- ✓ no adaptive question insertion possible at this point

At the bottom of the slide, there are two red question mark icons with small figures standing next to them.

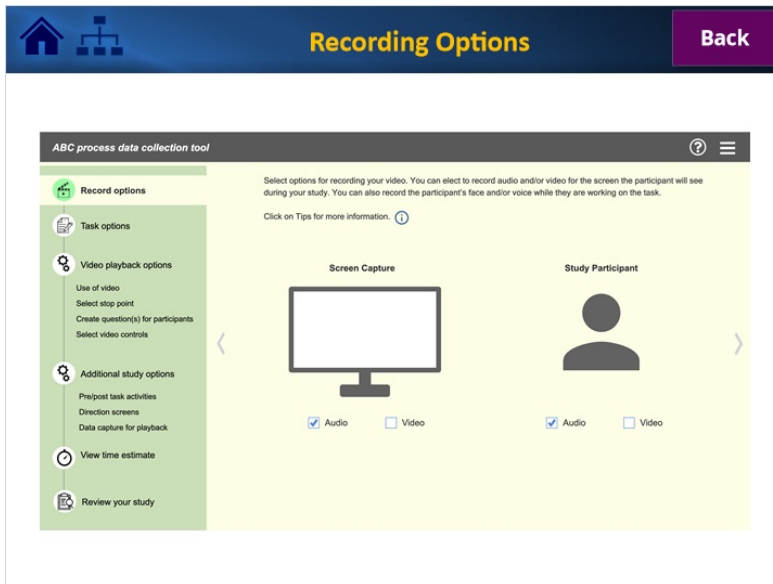
## 5.8 Experimental Design (II)



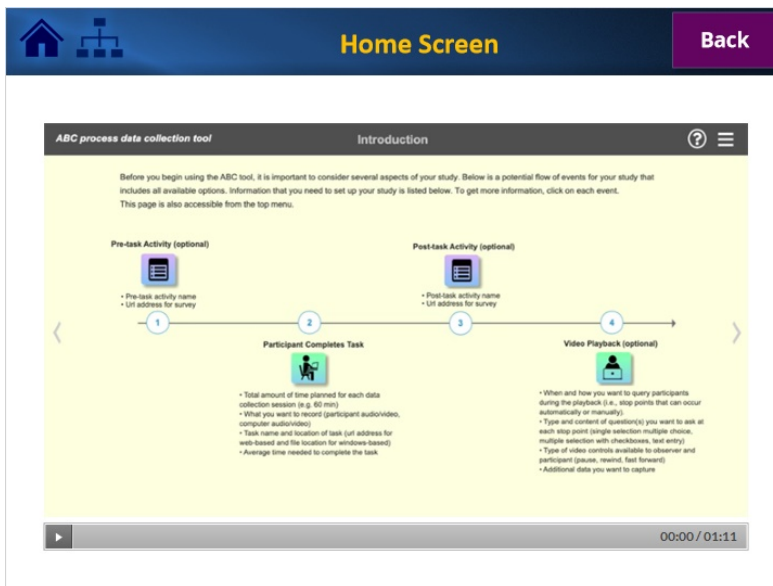
## 5.9 Tool Feature Selection





## 5.10 Recording Options



## 5.11 Main Screen



## 5.12 Playback Configuration



Playback Configuration

Back

ABC process data collection tool

Record options

Task options

Video playback options

Use of video

Select stop point

Create question(s) for participants

Select video controls

Additional study options

Prepost task activities

Direction screens

Data capture for playback

View time estimate

Review your study



You have selected that you want to use the video as part of your study. Next, you will decide how you want the video to stop during the playback.

☐ The observer or participant can only stop the video manually

☐ Set up automatic stops. You can allow the participant to stop the video manually as well.

00:00 / 02:08

## 5.13 Question Configuration



Question Configuration

Back

ABC process data collection tool

Record options

Task options

Video playback options

Use of video

Select stop point

Create question(s) for participants

Select video controls

Additional study options

Prepost task activities

Direction screens

Data capture for playback

View time estimate

Review your study

You selected stopping at specific locations in the video. You have also chosen the option to manually stop the video. Now you can create questions you want to ask the participants and link them to stop points. Each stop point can have as many as 4 questions.

Automatic Stop

| Video Timestamp | Question # |
|-----------------|------------|
| 00:30           | 1 2        |
| 1:00            |            |
| 2:00            |            |
| 3:00            |            |
| 4:00            |            |
| 5:00            |            |

Manual Stop

Question Pool

Drag questions and place them to the boxes on the left.

1 Are you motivated to do your best currently?

2 How do you currently feel?

3 Why did you stop the video now? What is going on?

4 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

5 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

6 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

7 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

8 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

9 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

10 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

11 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

12 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod b

Add a Question

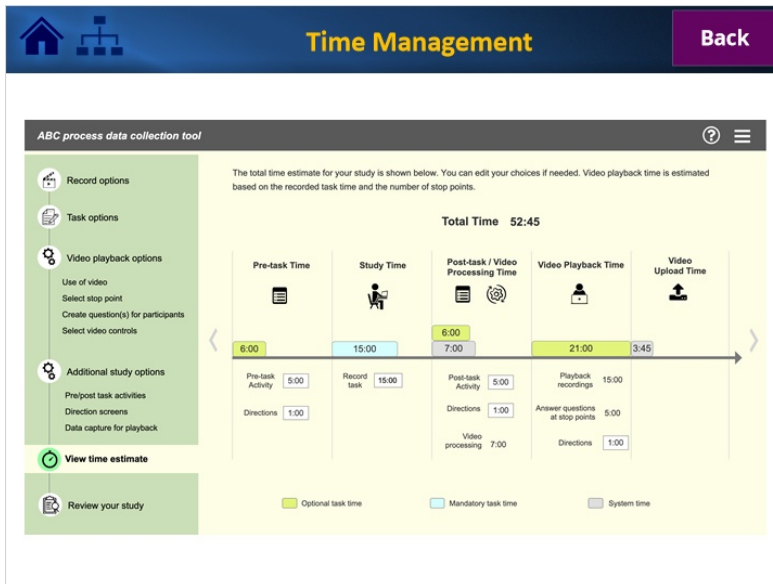
00:00 / 02:37

DM12 SLIDES (Verbal Reports, Version 1.1)

72 /

79

## 5.14 Time Management





## 6. Section 5: ABC Tool Activity

### 6.1 Cover: Section 5



## 6.2 Activity Overview




### Activity Overview



- **4 video clips** of participant playing the “Awkward Annie” game
- game designed to assess **English language pragmatics** in the workplace
- participants take the **role of Annie** in a virtual workplace

- participants engage in **conversations with coworkers** that differ by:
  - (a) relative power
  - (b) relative familiarity
- each conversation has **3 turns with 3 options**
- virtual **avatars react** via text, facial expressions, and gestures

- **2 versions** of game with **different goals**:
  - (a) awkward conversations
  - (b) polite conversations



## 6.3 Activity Overview



### Activity Overview

Watch clips and generate possible questions considering factors such as:

- (a) task features
- (b) participant cognition
- (c) participant response processes
- (d) participant experience

Video 1

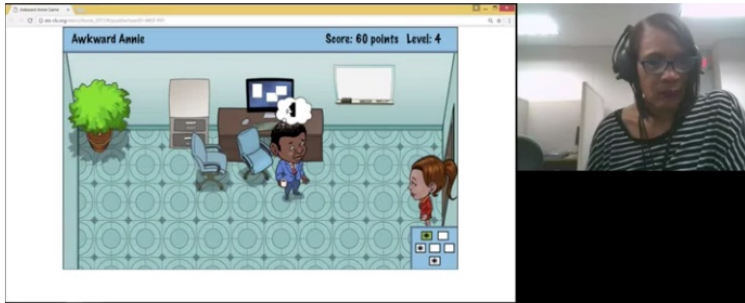
Video 2

Video 3

Video 4



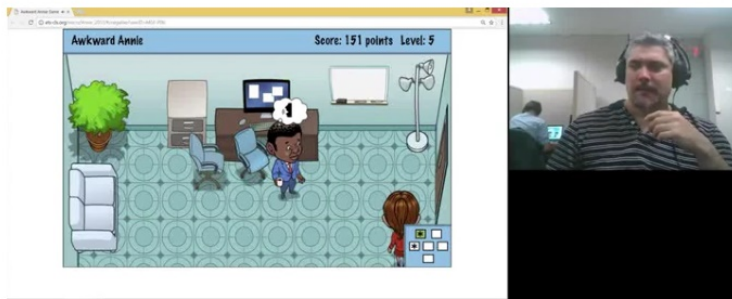
## Video 1 (Slide Layer)



Click on the video, which will open in a new browser window



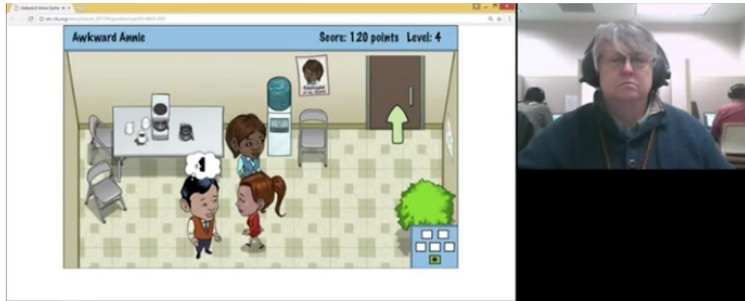
## Video 2 (Slide Layer)



Click on the video, which will open in a new browser window



### Video 3 (Slide Layer)



Click on the video, which will open in a new browser window





### Video 4 (Slide Layer)



Click on the video, which will open in a new browser window



## 6.4 Reflection Question 1




### Reflection Question 1



Given the task and the recording channels available in this experiment, what research questions would you ask?

**Please write down at least two research questions.**


Consider the different options for data collection that are afforded by the ABC Tool.



## 6.5 Reflection Question 2



### Reflection Question 2

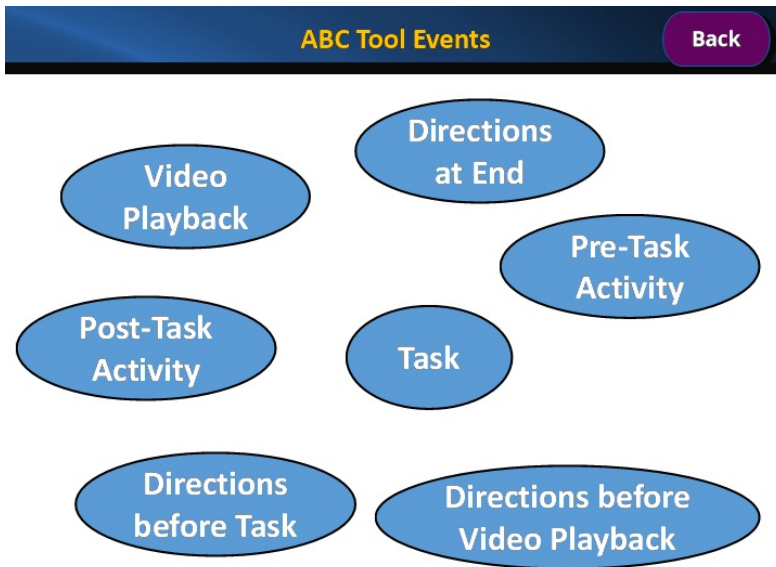


Given these research questions, what would your experiment look like utilizing the ABC Tool?

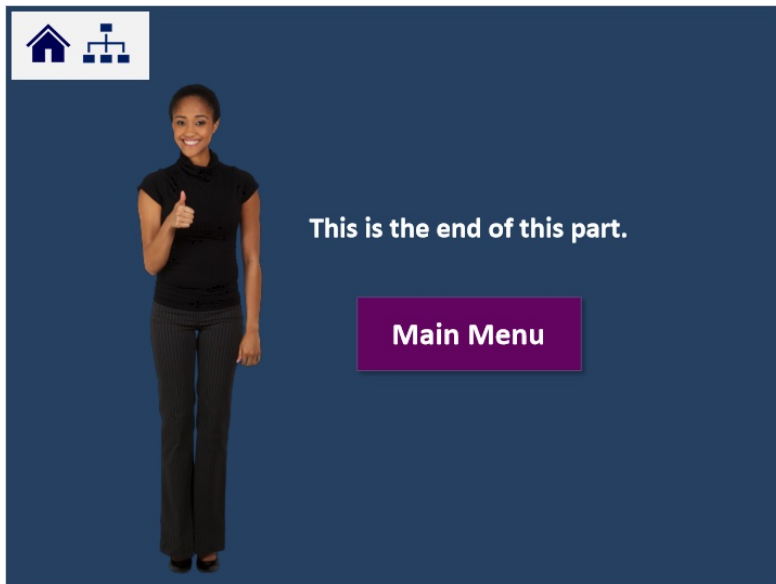
- ✓ Identify those ABC Tool events that you would like to include (click the button below for a visual reminder.)
- ✓ Next, write down a brief rationale for the choice of each event.
- ✓ Finally, discuss your choices with a colleague if you like.

**ABC Tool Events**

## ABC Tool Events (Slide Layer)



## 6.6 Bookend: Section 5



## 6.7 Module Cover

