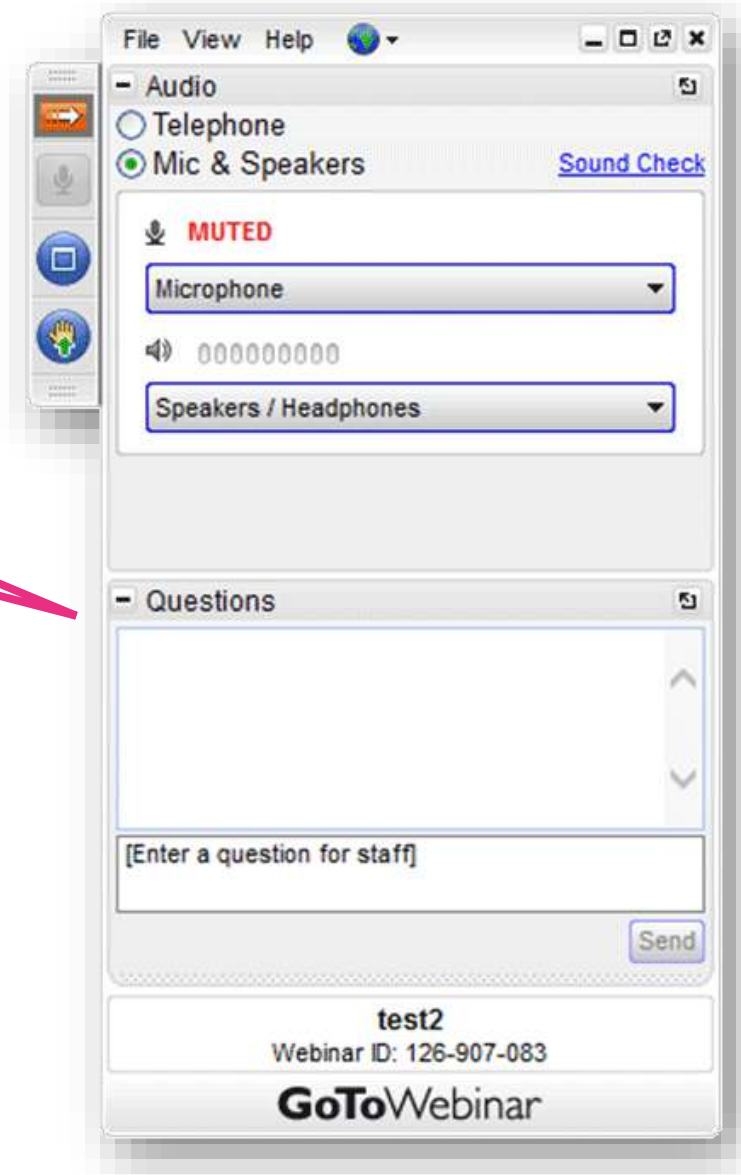# Item Writing: Tips & Techniques for Writing Good Questions

John Kleeman, *Questionmark*

To ask questions,
use the "Questions"
feature

**Watch for an email after the webinar:**
- Download slides (PDF)
- View a recording
- Answer a survey

About Questionmark

Background

- Founded in 1988

- Assessment solutions to measure knowledge, skills, abilities and attitudes securely for certification, regulatory compliance, workforce learning, sales-force readiness and higher education

- ISO/IEC 27001 Certified (Learn more: www.questionmark.com/trust)
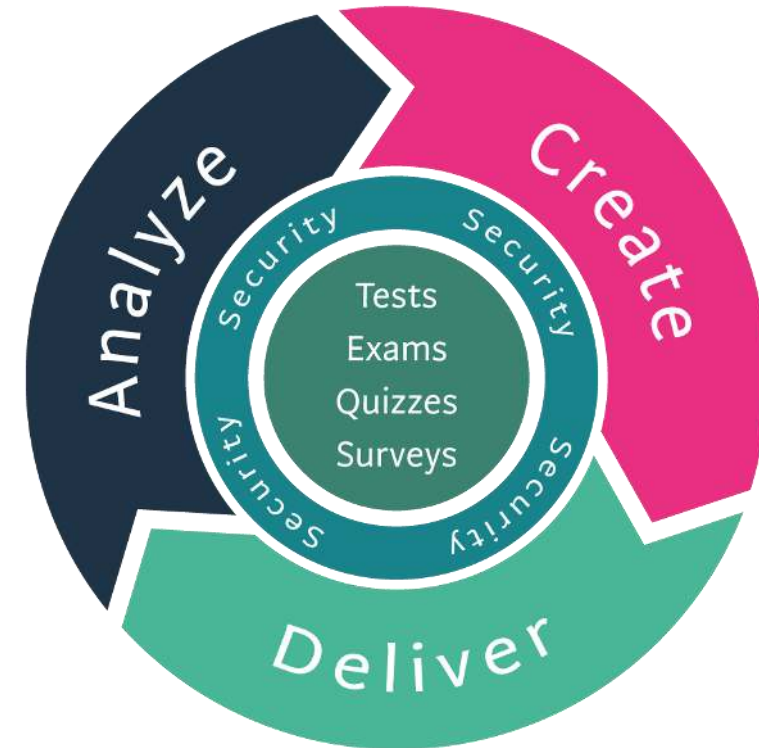
- *Questionmark OnDemand*
- *Questionmark OnDemand for Government*
- *Questionmark OnPremise*

Presenter:

- John Kleeman MA MBCS CEng CIPP/E

- Founder of Questionmark

- EVP External Relations, Learnosity and Questionmark

- 2021 Chairperson of Association of Test Publishers

- 30 years of experience in the assessment industry

Thanks to Brian McNamara and others at Questionmark for much of webinar content
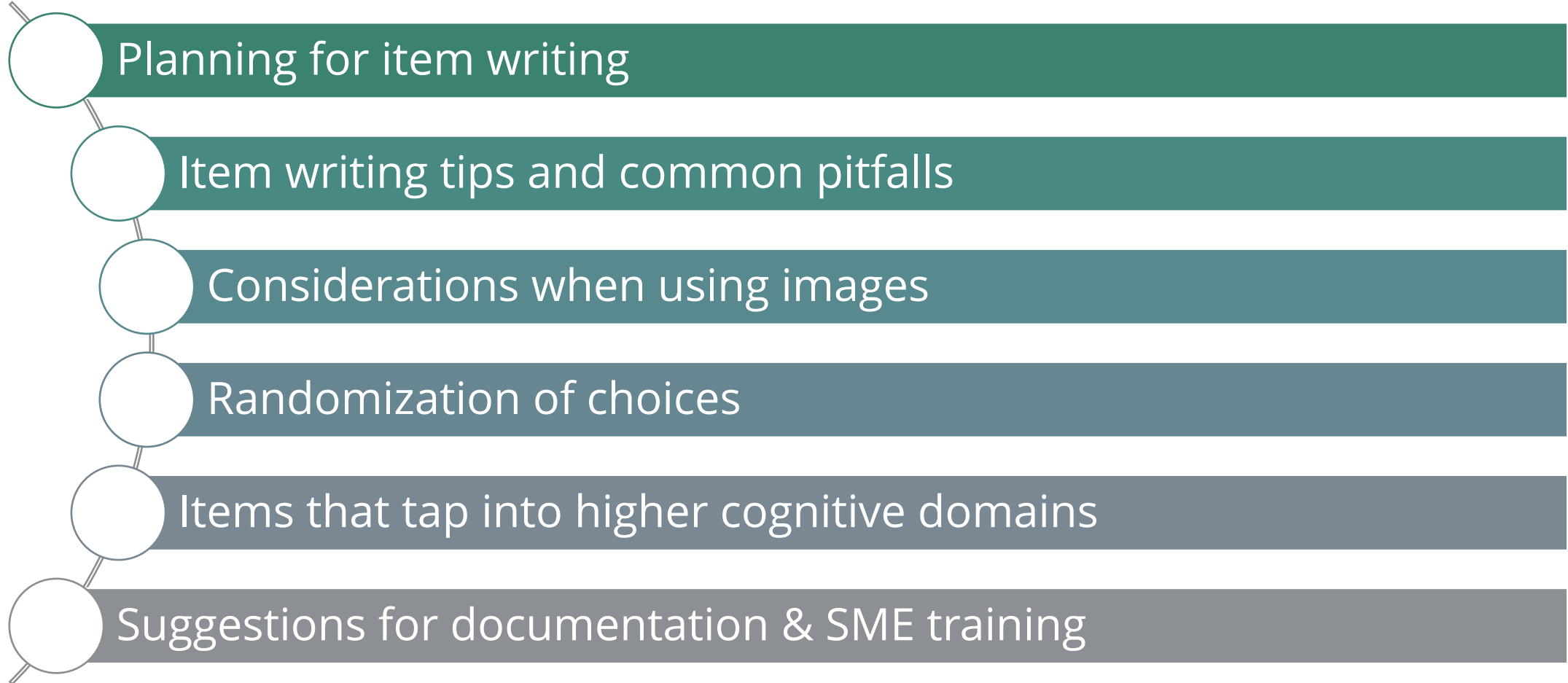
Quick Poll ☑

What brings you here today?

- I'm new to item writing
- I'm experienced in item writing but would like a refresher
- I supervise and/or coach other people who write test items
- I wanted to learn more about Questionmark

What we'll cover today...

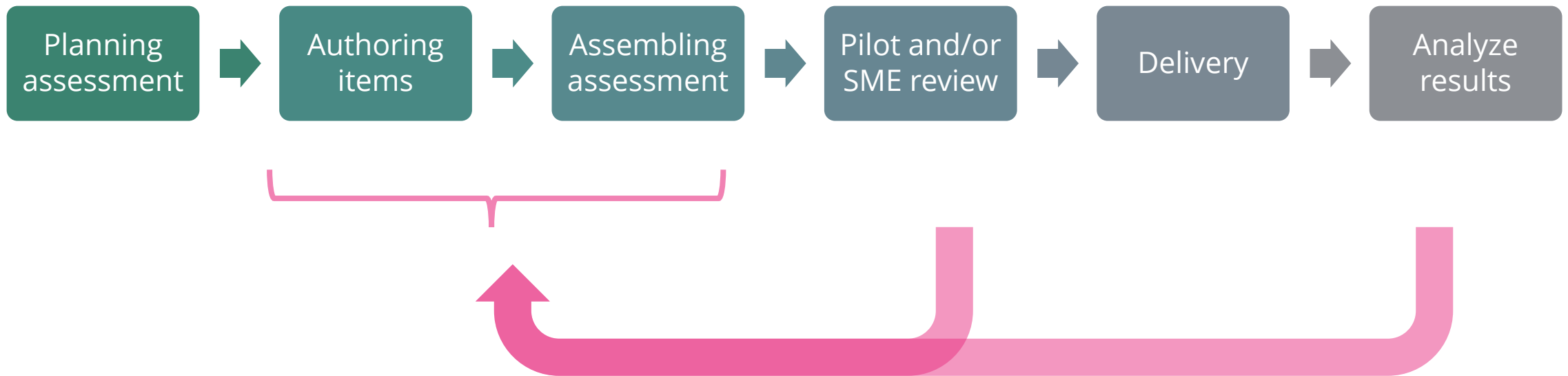- Planning for item writing
- Item writing tips and common pitfalls
- Considerations when using images
- Randomization of choices
- Items that tap into higher cognitive domains
- Suggestions for documentation & SME training

Planning for item writing

Item Writing within Assessment Development Process

Planning assessment → Authoring items → Assembling assessment → Pilot and/or SME review → Delivery → Analyze results

Planning for Item Writing: Blueprints

- Document content or behaviors that are being assessed and relative weights for each area (APA, NCME, & AERA 2014)

- Can also include:
  - Specified item formats
  - Target difficulty levels
  - Cognitive levels (Raymond & Neustel, 2006)
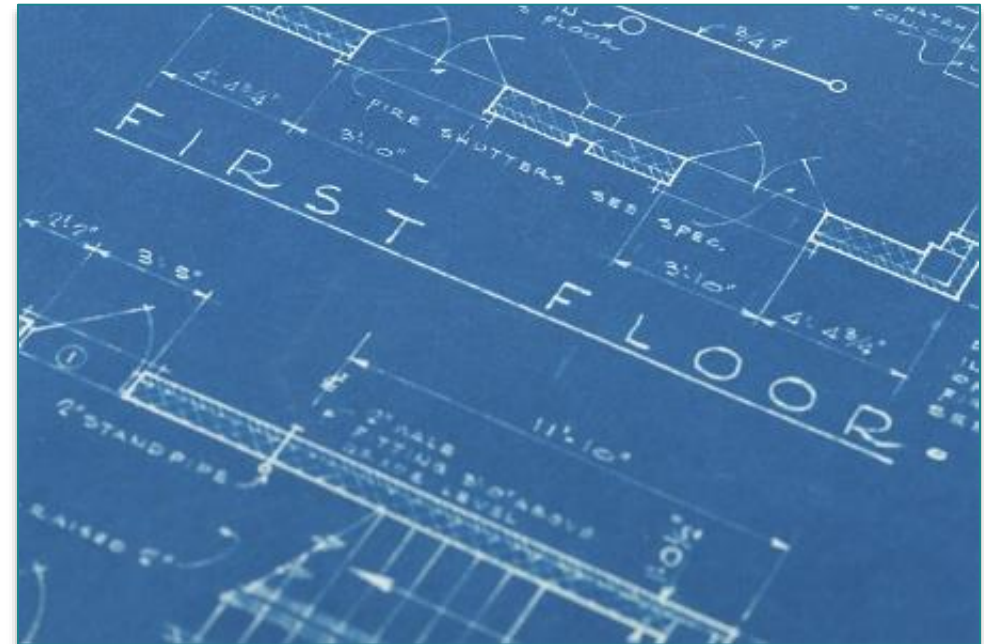
Simple Example of Assessment Blueprint

Electrical Skills Test - Blueprint	% of test	# items	Skill level	Criticality
Section 1: AC Motors	45%	**34**		
1.a - Lorem ipsum dolor		21	3	3
1.b - Duis aute irure		8	3	3
1.c - Excepteur sint occaecat		5	2	3
Section 2: High-voltage switchgear	25%	**19**		
2.a - Neque porro quisquam		12	2	2
2.b - Modi tempora incidunt		7	3	2
Section 3: Synchronous Motors	30%	**23**		
3.a - Totam rem aperiam		11	2	2
3.b - Eaque ipsa quae		12	2	3

Approaches to developing test blueprints

Process-Based Blueprints

- Define processes to be assessed, such as:
  - Cognitive processes
  - Processes related to job performance
- Process-based blueprints are **built around verbs**, often in the form of learning objectives (Raymond & Neustel, 2006)
  - "participant is able to *analyze* the difference between. . ."
  - "participant can *operate* the equipment needed to. . ."
- Provides guidance on types of behaviors items should be designed to elicit

Content-Based Blueprints

- Define content to be assessed, such as:
  - Curricular or training content
  - Problems often encountered on the job
- Content-based blueprints often **are based on nouns** that identify topics (Raymond & Neustel, 2006)
  - "Algebra"
  - "Client Support"
- Easy to communicate content of the assessment, and easily mapped to curriculum

More Complex Blueprint : Content and Process Together

Cognitive Levels
(Questionmark meta tags)

Content	Weight	# Items	Remember		Understand		Apply		Analyze		Evaluate		Create	
			Cognitive Level											
Ingredients	50%	50	to list three ingredients in a PB&J		to explain the flavor and texture profiles of ingredients in a PB&J		to find the required ingredients at a grocery store		to differentiate between unlabeled jars of crunchy PB and creamy PB		to contrast the ingredients used in a PB&J and a Velvet Elvis			
			Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items
			25%	12-13	25%	12-13	25%	12-13	15%	7-8	10%	5		
Assembly	30%	30	to list the order of ingredients in a completed PB&J		to explain potential problems that arise when assembly is not done properly		to prepare a PB&J, given the ingredients and kitchen tools		to identify PB&Js that have been assembled incorrectly		to explain what steps might have been done incorrectly, when given an incorrectly assembled PB& to review			
			Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items
			40%	12	30%	9	10%	3	10%	3	10%	3		
Presentation	20%	20	to describe principles of proper PB&J presentation		to predict if a customer will be able to eat a PB&J without getting sticky fingers, given a presentation scenario		to remove crusts and cut PB&J at various angles							
			Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items	Weight	# Items
			50%	10	20%	4	30%	6						

Content
(Questionmark Topics)

Weight on assessment

Processes
(i.e. Learning Objectives)

Job Task Analysis Surveys to guide item development

2 of 3
What is your role in the organization? [▾]

3 of 3
Answer questions about nursing.

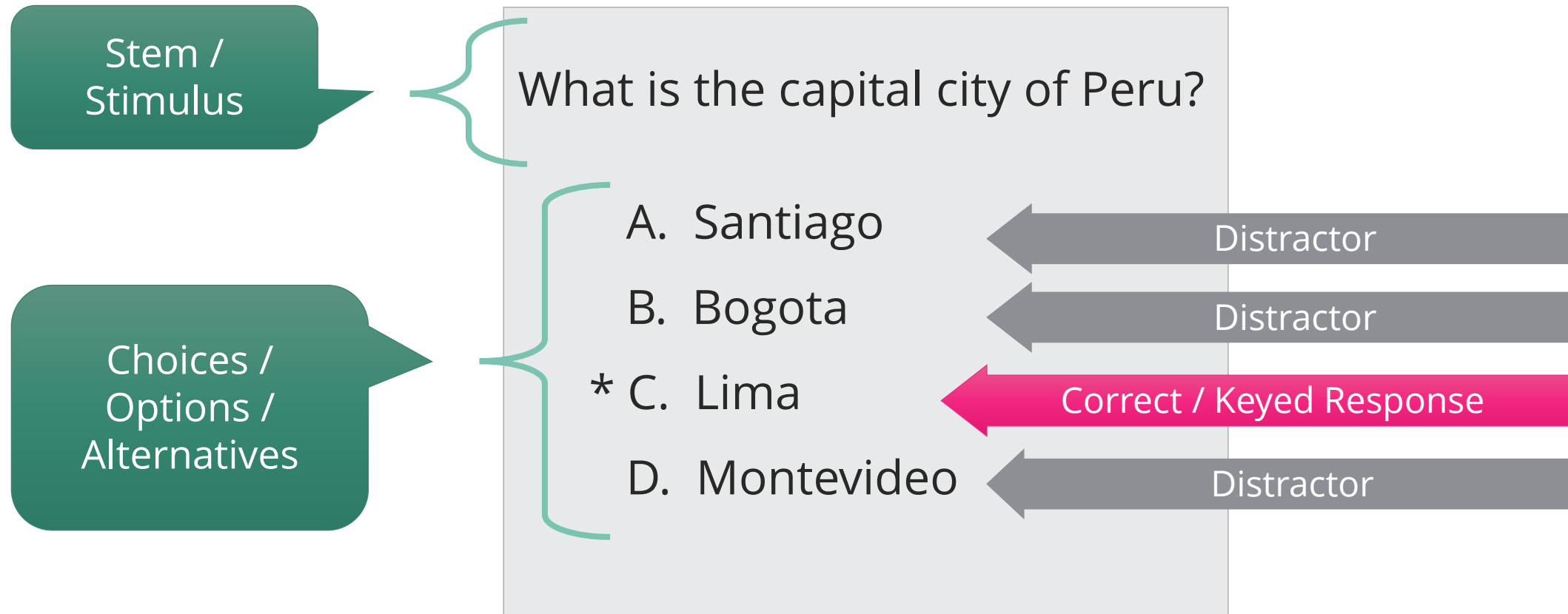|	Applicability			Difficulty					Not Important	Somewhat Important
	Do Task	Supervise Task	N/A	Very Easy	Easy	Neither Easy or Difficult	Difficult	Very Difficult		
Administering medication	○	○	○	○	○	○	○	○	○	○
Assessing patients	○	○	○	○	○	○	○	○	○	
Assisting patient	○	○	○	○	○	○	○	○	○	
Communicating with family members	○	○	○	○	○	○	○	○	○	
Cleaning surgical area	○	○	○	○	○	○	○	○	○	
Showing empathy	○	○	○	○	○	○	○	○	○	

Survey SMEs about key tasks:

How Difficult?
How Important?
How Frequent?
How Critical?

- Job Task Analysis to determine and **validate** content of test
- Test Blueprint built from JTA, which drives test item development.

Terminology/Basic Structure of a Multiple-Choice Item

Stem / Stimulus

What is the capital city of Peru?

Choices / Options / Alternatives

A. Santiago — Distractor

B. Bogota — Distractor

* C. Lima — Correct / Keyed Response

D. Montevideo — Distractor

Item feedback

Which of the following hazards must the forklift operator consider when arranging a load?

- ○ Loss of steering control
- ● Tip-over of the vehicle
- ○ Collision
- ○ Falling load
- ○ All of the above

0 out of 1

The correct answer is "all of the above"

Reminder on Required Practices:

- Do not exceed the capacity of the forklift that appears on the forklift's data plate.
- Minimize the distance from the front wheels to the load center.
- Load a large rectangular box width wise across the forks of the truck as in Figure 2.
- The heaviest weight should be loaded as close to the masts as possible.

4500 lbs. @ 24" LC
24"

4500 lbs. @ 30" LC
30"

Review full instructions here.
https://www.osha.gov/SLTC/etools/pit/operations/loadcomposition.html

Question showing participant's answer

Corrective feedback

Item writing tips and common pitfalls

Yes or No:
When loading your snarkleblaster, should you beedle the diggle or zix the frondle?

- o Yes
- o No

Which of the following choices best completes this sentence?

A gorilla's favorite treat is a ___ .

 a. Apple

 b. Banana

 c. Orange

 d. Eggplant

question mark

Quick Poll ☑

What problem do you spot in this question?

- Uses negatives or other poor language
- Grammatical cues give away the answer
- It's asking more than one question
- Distractors are not plausible

Arrange the parents of the following children with Down's syndrome in order of highest to lowest risk of recurrence. Assume that the maternal age in all cases is 22 years and that a subsequent pregnancy occurs within 5 years. The karyotypes of the daughters are:

I: 46, XX, -14, +T(14q21q) pat
II: 46, XX, -14, +T(14q21q) de novo
III: 46, XX, -14, +T(14q21q) mut
IV: 46, XX, -21, +T(14q21q) pat
V: 47, XX, -21, +T (21q21q) (parents not karyotyped)

A: III, IV, I, V, II
B: IV, III, V, I, II
C: III, I, IV, V, II
D: IV, III, I, V, II
E: III, IV, I, II, V

Avoid unnecessarily complexity

Source: National Board of Medical Examiners "Constructing Written Test Questions for the Basic and Clinical Sciences"

Which factors impact a presidential election?

Select all that apply.

- ☐ Previous president's performance
- ☐ Weather
- ☐ Congress' performance rating
- ☐ Local economy
- ☐ National unemployment rate

Consider alternate question types

All of the Above (AOTA) &
None of the Above (NOTA)

Risks of AOTA and NOTA

Issues with "All of the Above"	Issues with "None of the Above"
• Recognizing at least two distractors as correct allows participant to logically assume AOTA is the appropriate answer • Recognizing at least one of the distractors as incorrect eliminates the AOTA option	• Defensibility: If there are any issues with the keyed correct response, then NOTA could arguably be a correct response

Well Written Questions: **Stimulus**

- Clear / Carefully worded
- Concise (avoid window dressing)
- Avoid jargon
- Positive phrasing
- Specific focus
- Ask a single question
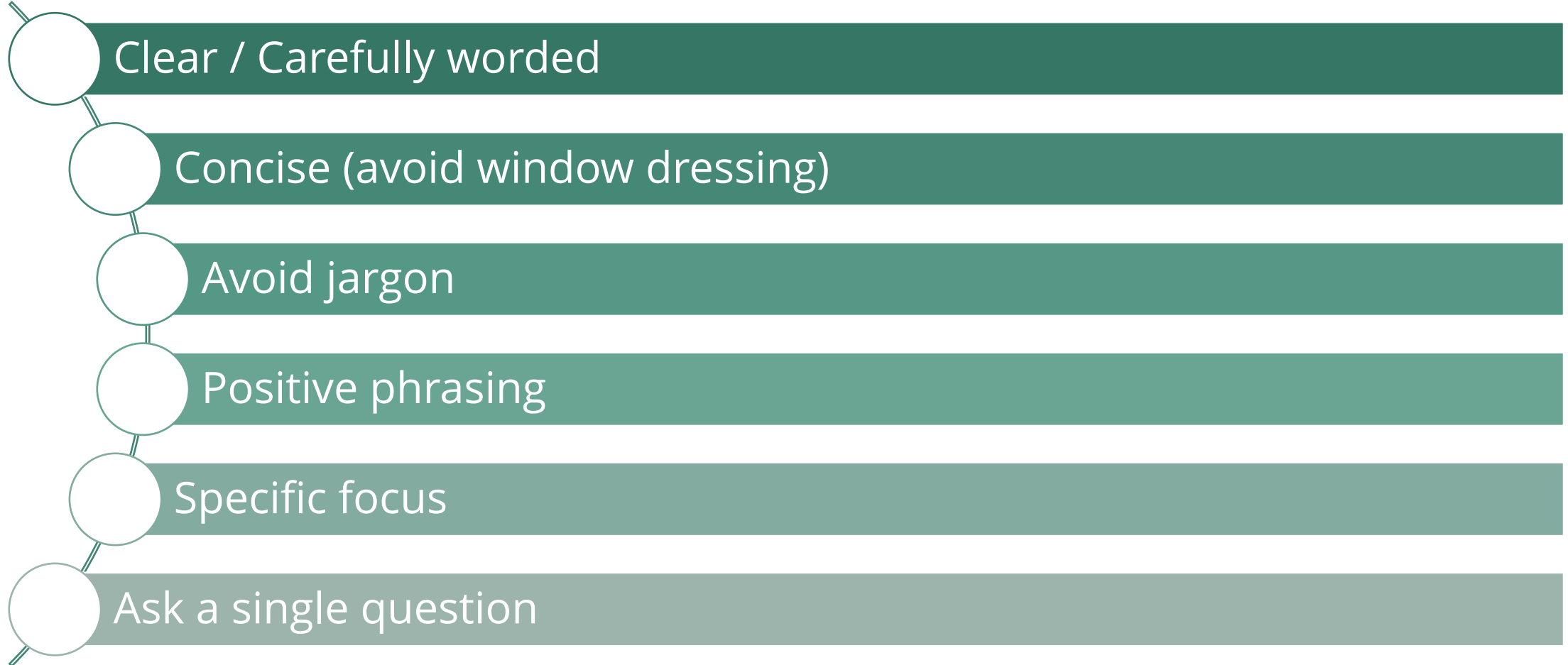
Well Written Questions: **Choices**

- Only one correct answer (Multiple choice)
- Plausible distractors
- Distractors featuring common misconceptions
- Roughly same length
- As short as possible
- Avoid keywords from within stimulus
- Avoid grammatical cues

Tips for avoiding bias / stereotyping

- Use neutral terms (sales agent instead of salesman)
- Standard, formal language
- Avoid references to race, ethnicity, gender, age, etc.
- Avoid idioms and colloquialisms
- Avoid obscure language, ambiguous acronyms, etc.

Randomizing/Shuffling of Choices

Randomizing/Shuffling of Options

- Easy to accomplish with many platforms

- Can be useful to:

  o Mitigate cheating/copying

  o Keep choice ordering objective
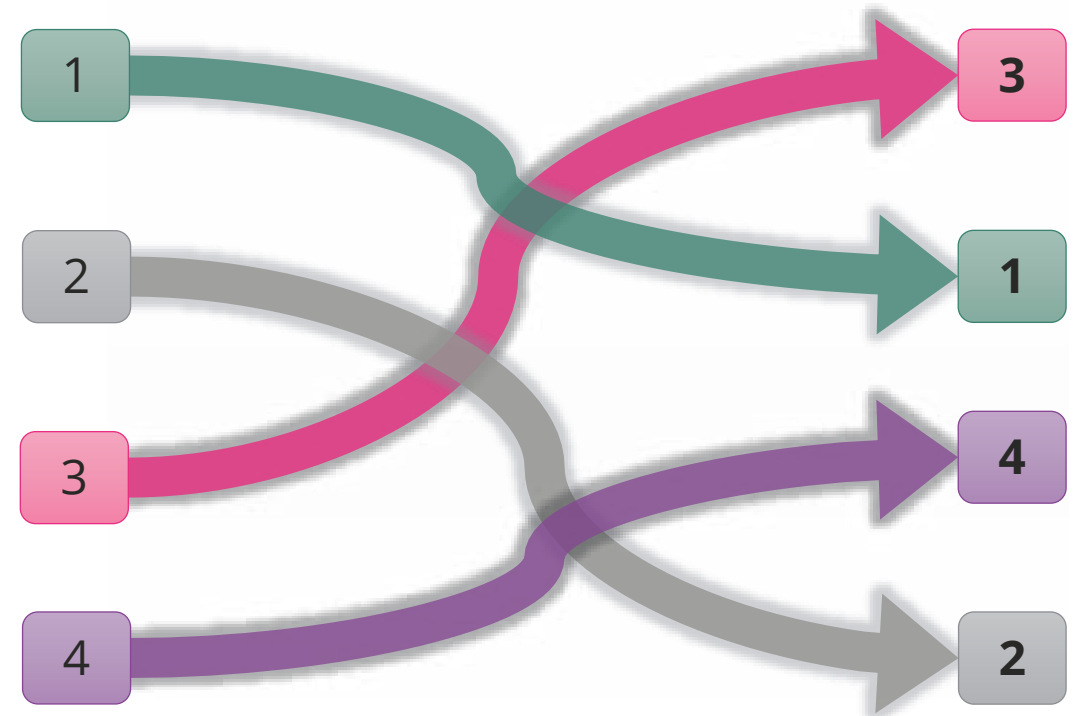
Can randomization interfere with measurement?

Logical or Numerical Order	Randomized Order

Which of the following choices is the correct answer to this equation?

-130 x -35 + 120 = _____

    A. -4550

    B. -4430

    C.   4550

* D.   4670

Which of the following choices is the correct answer to this equation?

-130 x -35 + 120 = _____

    A. -4430

* B.   4670

    C. -4550

    D.   4550

Considerations

- When using AOTA or NOTA

Shuffling option

All
No shuffling
All
All but last choice
All but last two choices

- Is randomized order of choices confusing to the participant

Study involving two mathematics tests identical except for:

- Version 1: Choices in logical (ascending/descending) order
- Version 2: Randomized choices

Did randomization impact item response stats?

- **Item difficulty:**   Unaffected
- **Time to answer:**   Unaffected
- **Discrimination:**   **Increased**

Study concludes that random ordering of numerical options:

- Is not an obstacle for upper-ability examinees
- May be an obstacle for lower-ability examines

"Numerical Answer Options: Logical or Random Order?" - Renee M. Huntley and Catherine J. Welch, 1993 (https://files.eric.ed.gov/fulltext/ED358136.pdf)

Images, figures and diagrams

Consider this question:

You are operating a forklift rated for a 6000-pound capacity at a 24-inch load center.

What is the maximum load capacity of the forklift to safely move a crate with a 36-inch load center?

A. 3000 pounds

B. 4000 pounds

C. 6000 pounds

D. 9000 pounds

Answering this question requires participant to understand:

- Terminology (e.g. "load center")

- How to calculate maximum load weight:
  - Divide truck's load center by crate's load center
 - Multiply by truck's maximum capacity.

Correct answer:

B. 4000 pounds

*(24 / 36) * 6000 = 4000 pounds*

Will graphic with stimulus impact measurement?

You are operating a forklift rated for a 6000-pound capacity at a 24-inch load center.

What is the maximum load capacity of the forklift to safely move a crate with a 36-inch load center?

A. 3000 pounds

B. 4000 pounds

C. 6000 pounds

D. 9000 pounds

Max. 6000 lb. @ 24"

24 in.

36 in.
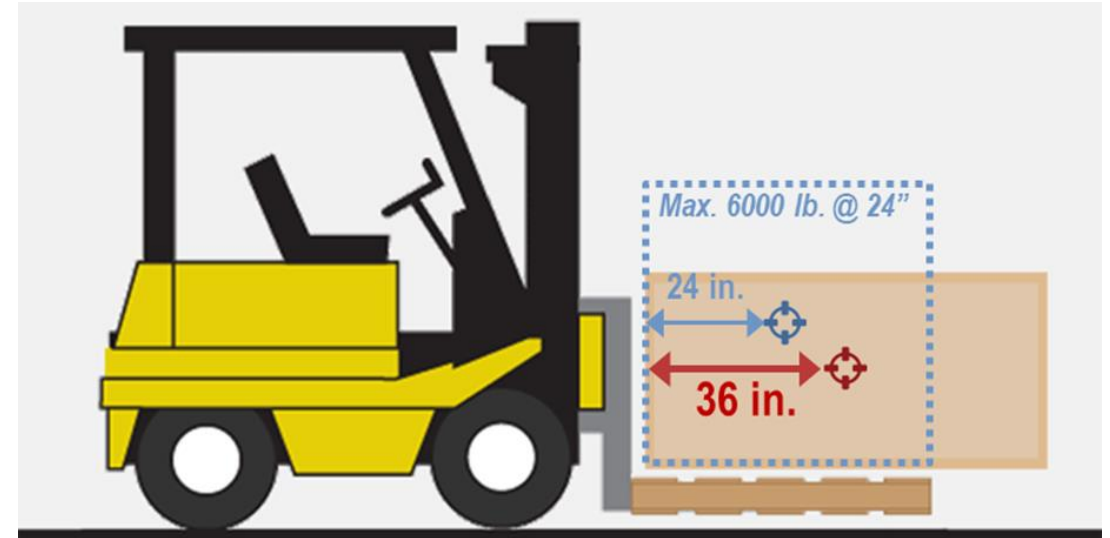
- Does the figure give away the answer?
  - May make easier to guess... addressable with a numeric question
- Or does it increase fidelity?

Guidance on using images, illustrations and diagrams
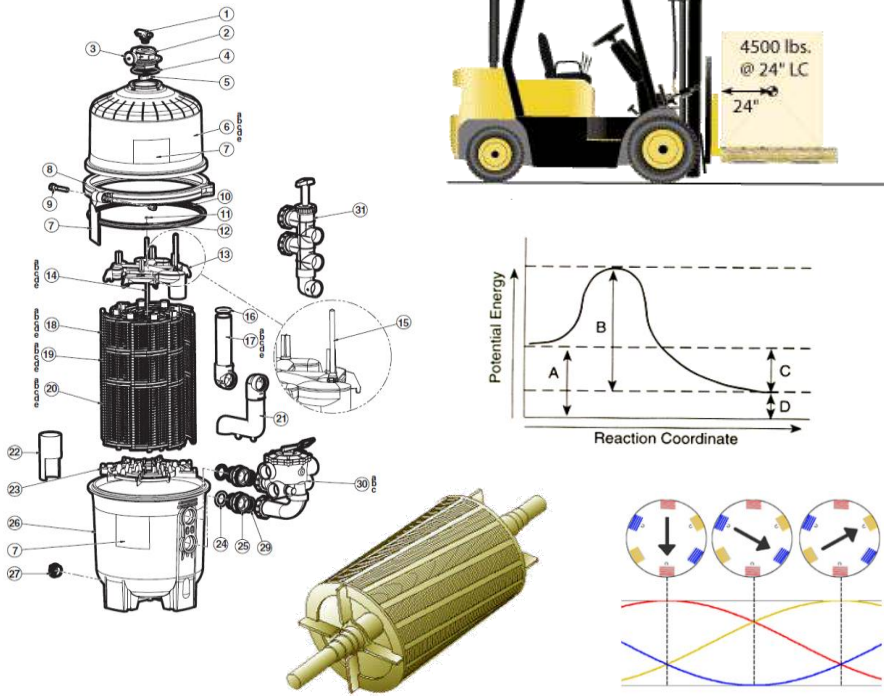
Use when appropriate and/or potential to enhances fidelity

Show figure with test item

Ensure clarity, appropriate size and perspective

Consider and address accessibility

Writing questions that measure "above" the knowledge/memorization level

Quick Poll

Can a multiple choice question be used to test "above knowledge"?

- Yes
- No
- Don't know

"Recall" is only small part of job requirements

- Blooms taxonomy describes cognitive skills in levels

- Most job roles require many levels of taxonomy

- **If your assessments focus <u>only</u> on knowledge recall, may well not test job competence**

Bloom's Taxonomy (revised version)

Create

Evaluate

Analyze

Apply

Understand

Remember / Recall
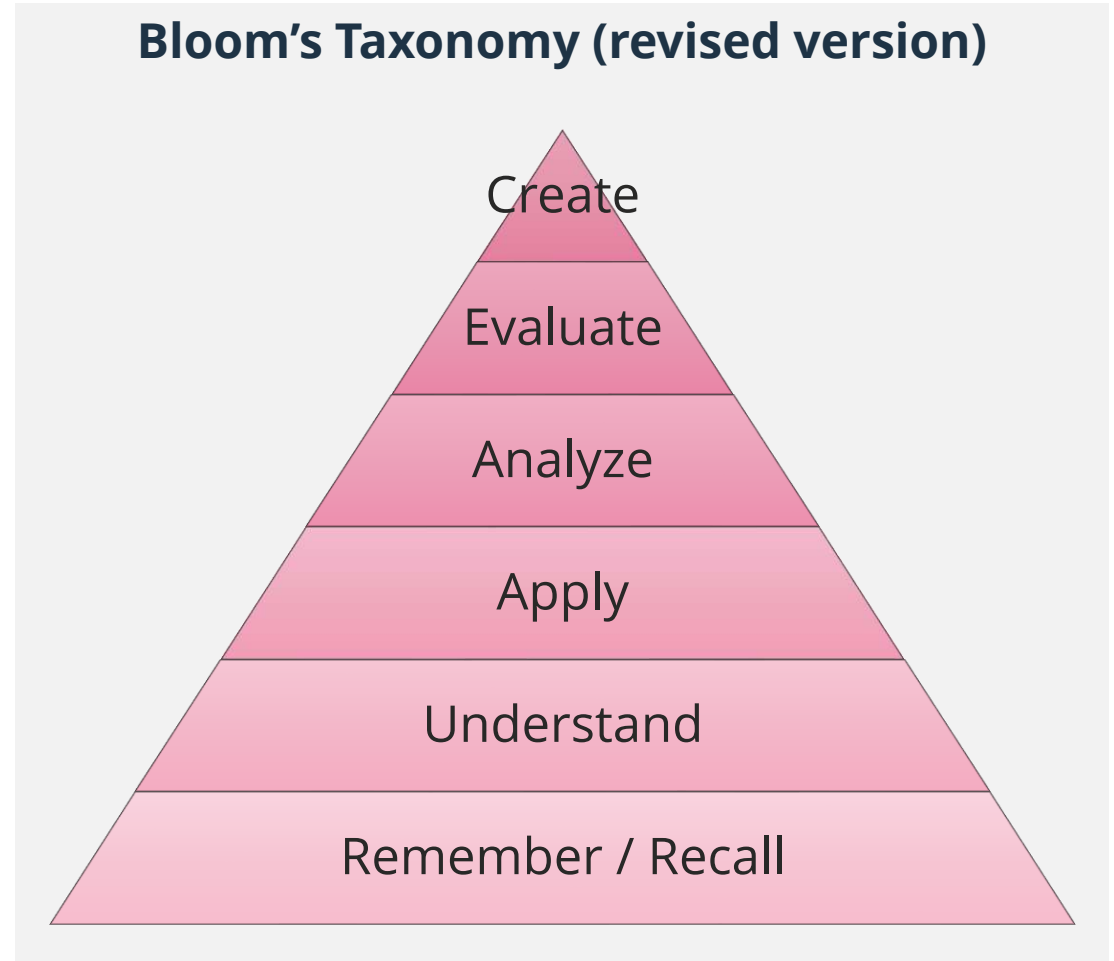
"Recall" example

What does a yellow traffic light mean?

a) Stop
b) Go
c) Caution
d) Look behind you

"Analyze/Evaluate" example

You are giving a close friend a ride to work and because of traffic you are running 15 minutes late for an important appointment. You are driving toward an intersection with traffic lights on it. You can see the intersecting road to some extent and there seem to be no cars on it. The light turns to yellow as you approach. The car behind you sounds their horn and seems to be accelerating in the expectation that you are going to cross the intersection. Your friend tells you to put your foot down as you are running late and there are no cars visible on the intersection. What should you do?

a) The road appears to be clear so speed up and cross the intersection to exit the intersection as soon as possible

b) Stay at the same speed and continue to cross the intersection while keeping an eye out for crossing cars

c) The light is about to turn red so you should stop immediately and ignore the car behind you as they will need to stop as well

d) Put your hazard lights on to signal to the car behind you and stop at the intersection if the lights facing you turn red

e) Ignore your friend's irritation and move across to the side of the road to let the other car pass

...a more realistic example

Recall	Analyze/Evaluate

Recall

Which of the following assessment findings is characteristic of a client with Parkinson's disease?

    A. Night blindness
    B. Pain in lower extremities
    C. Shuffling gait
    D. Incontinence

Source: Writing Multiple-Choice Test Items that Promote and Measure Critical Thinking, Morrison and Free. Journal of Nursing Education, Jan 2001

Analyze/Evaluate

The nurse is making a home visit to a 75-year old male client who has had Parkinson's disease for the past five years. Which finding has the greatest implication for this client's care?

   A. The client's wife tells the nurse that the grandchildren have not been to visit for over a month.
    B. The nurse notes that there are numerous throw rugs throughout the client's home.
    C. The client has a towel wrapped round his neck that the wife uses to wipe her husband's face.
    D. The client is sitting in an am chair, and the nurse notes that he is gripping the arms of his chair.

Advice from the experts

William Coscarelli & Sharon Shrock
(in *Criterion-Referenced Test Development*)

- Knowledge (recall) level items are the easiest to write
- Developing tests that truly reflect OJT performance requires writing items focused on higher cognitive levels

In general, the single most useful improvement you can make in writing test items is to write them <u>above</u> the memorization level.

Create

Evaluate

Analyze

Apply

Understand

Remember / Recall
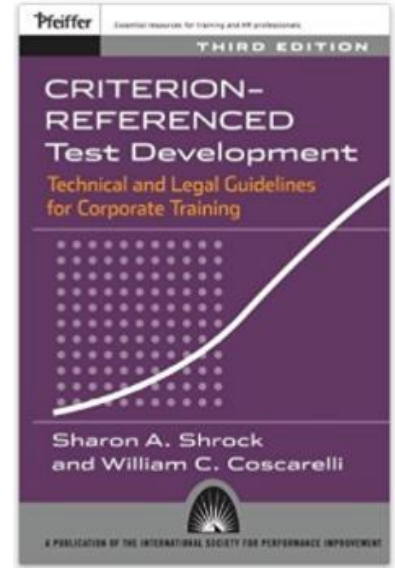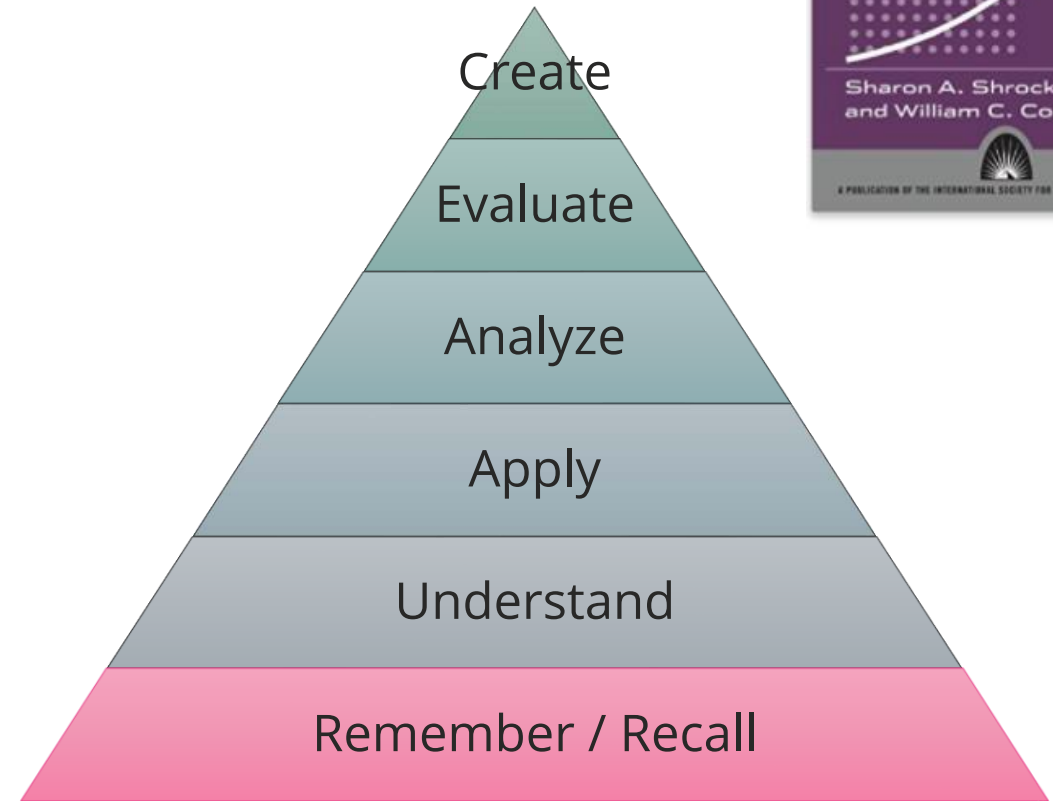
Writing conventional questions to assess beyond recall

Consider questions that require a participant to consider multiple facts (not just one fact) to answer

Include a scenario or rich context (e.g. video) in question to put participants closer to the work environment.

Include stimulus that requires analysis (e.g. charts and tables; images; etc.) for the participant to arrive at the correct answer

Consider the "representative verbs" correlating to the different levels of Bloom's taxonomy (see table on next slide)

Bloom's taxonomy with representative verbs and example behaviors:

Taxonomy Level	Representative Verbs	Sample behavior
Create	Organize, Design	Develop a new protocol for treating the cold
Evaluate	Critique, Summarize	Assess the effectiveness of that protocol
Analyze	Compare, Categorize	Compare and contrast progression of cold and flu, or Determine if a patient has a cold or the flu
Apply	Organize, Solve	Describe the standard process for determining if a patient has a cold or the flu
Understand	Distinguish, Match	Match symptoms with their associated ailments
Recall	Identify, Label	Identify three symptoms of a cold

Source: "Writing Multiple-Choice Questions for Higher-level Thinking," Mike Dickinson in *Learning Solutions,* 5-Dec-2011
https://www.learningsolutionsmag.com/articles/804/writing-multiple-choice-questions-for-higher-level-thinking?utm_campaign=lsmag&utm_medium=email&utm_source=lsm-news
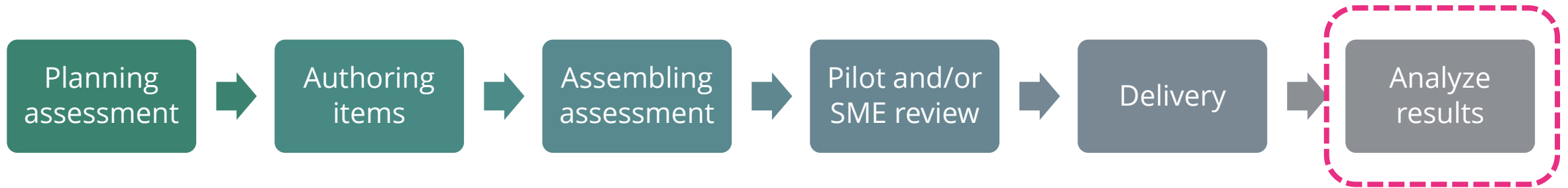
What is the **easternmost** state in the United States?

A. Alaska

B. Florida

C. Hawaii

D. Maine

Is this a good question?

"Option Analysis" of response data indicates a problem

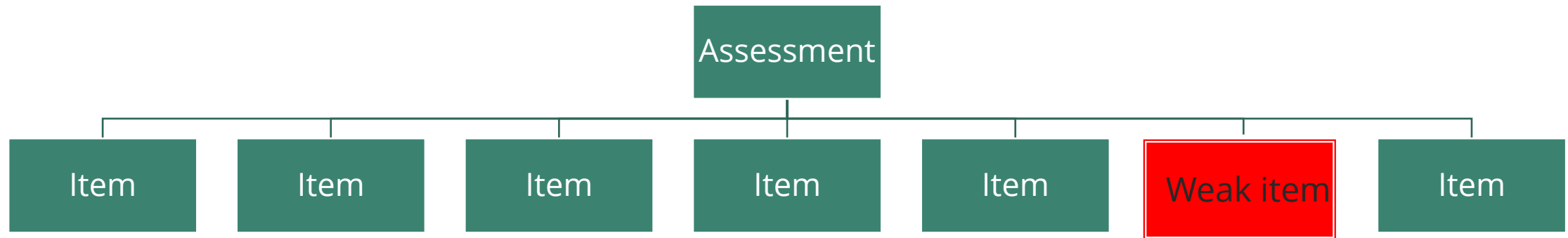43.3% of our top-performing participants selected "Alaska."

Question preview

What is the easternmost state in the United States?

○	Alaska
○	Florida
○	Hawaii
○	Maine

Answer option information		Number and percentage of participants achieving scores			
Outcome #	Answer option	All	Upper 27%	Middle 46%	Lower 27%
1	Alaska	278 (27.8%)	117 (43.3%)	91 (19.8%)	70 (25.9%)
2	Florida	146 (14.6%)	17 (6.3%)	63 (13.7%)	66 (24.4%)
3	Hawaii	158 (15.8%)	17 (6.3%)	71 (15.4%)	70 (25.9%)
✓ 4	Maine	418 (41.8%)	119 (44.1%)	235 (51.1%)	64 (23.7%)
5	No response	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total assessment mean score		51.8 %	70.5 %	53.5 %	30 %

Item statistics can identify poorly performing questions

```
                        ┌──────────────┐
                        │  Assessment  │
                        └──────┬───────┘
   ┌────────┬────────┬────────┼────────┬──────────┬────────┐
┌──┴──┐ ┌──┴──┐ ┌──┴──┐ ┌──┴──┐ ┌──┴──┐ ┌────┴────┐ ┌──┴──┐
│Item │ │Item │ │Item │ │Item │ │Item │ │Weak item│ │Item │
└─────┘ └─────┘ └─────┘ └─────┘ └─────┘ └─────────┘ └─────┘
```

- **Item Analysis** looks at the performance of each item

  o Are my questions too easy? Too difficult?
  o How effective are my distractors?
  o How well does the item differentiate between participants of different knowledge/skill levels?

- Removing/improving weak items makes the assessment more reliable and valid

Document all processes – especially for high-stakes

Reproducibility

- Follow the same steps and use the same resources if you need to maintain the assessment or build a new assessment in the future

Accountability

- Demonstrate that best-practices were followed in the event of a challenge or dispute
- Collect evidence needed for third party accreditation

Organization

- Keep track of assessments or content that are in different phases of development

Validity evidence

- Provide evidence to support validity arguments for the inferences of your assessment

Training Item Writers

- Policies, especially security

- Classifying items by content, cognitive levels

- Blueprints

- Intro item writing
  - Clear stems, plausible distractors, accurate keys
  - Writing to appropriate difficulty levels and cognitive levels
  - Appropriate uses of different item formats
  - Examples of good and bad items with opportunity to practice

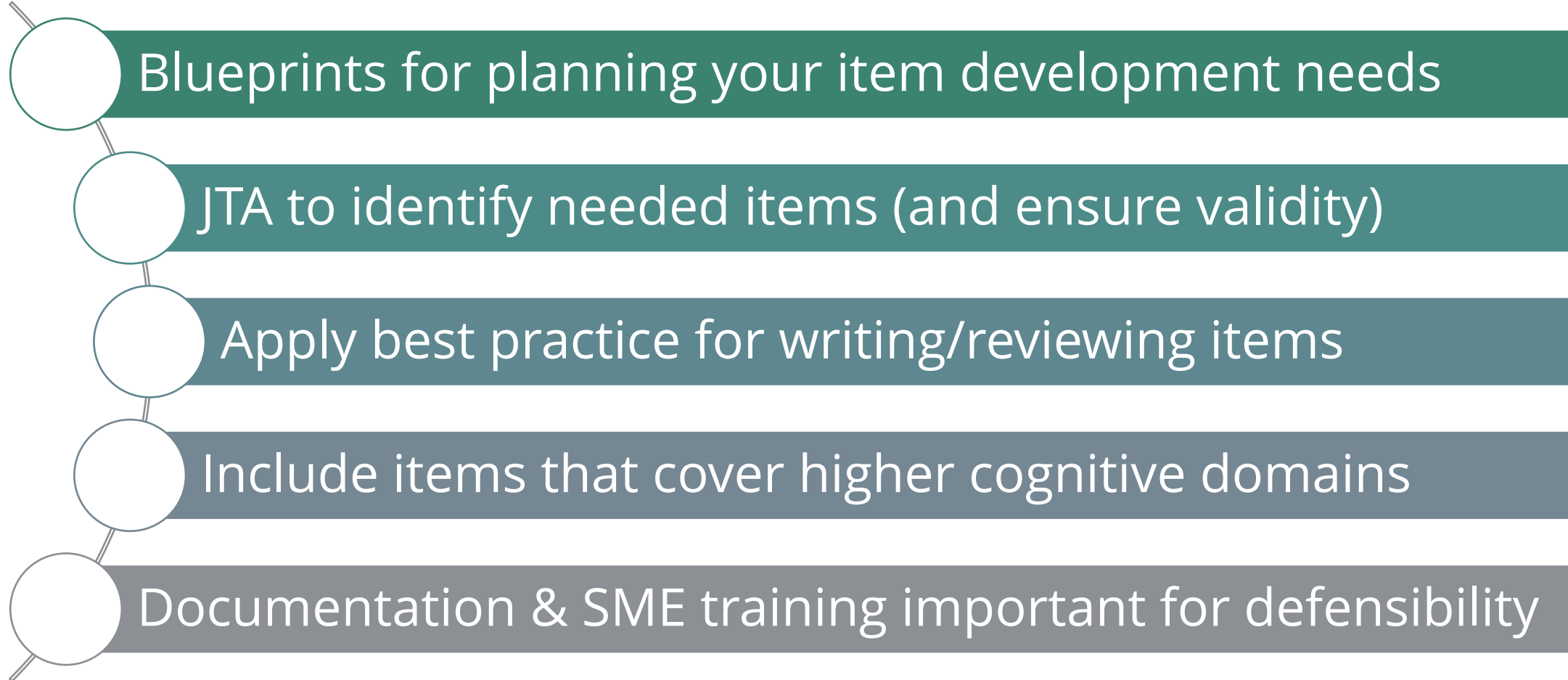- Style guides, scoring guides, rules for referencing items
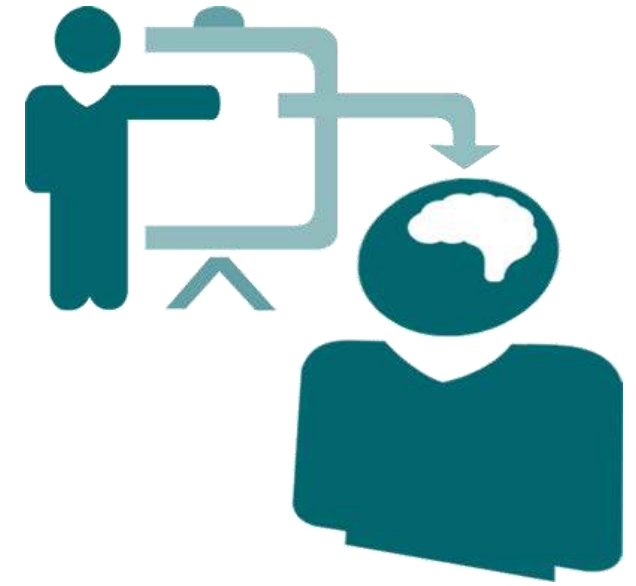
Examples of Style Guide Content

- Terms and Vocabulary
- Content format
  - Fonts/text sizes
  - Formatting of charges, tables, etc.
- Item format
  - Allowable items types
  - Scoring rules
  - # of distractors
  - How to order options
  - Use of images/diagrams
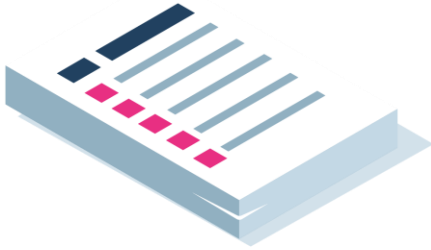 - Suggested templates
- Item format

Summary

Blueprints for planning your item development needs

JTA to identify needed items (and ensure validity)

Apply best practice for writing/reviewing items

Include items that cover higher cognitive domains

Documentation & SME training important for defensibility

White papers, infographics, reports, eBooks and more!

www.questionmark.com/resources
www.questionmark.com/resources/blog

Useful resources:

White Papers:
- Assessing for Situational Judgment
- Assessment Results You Can Trust
- Defensibility and Legal Certainty for Tests and Exams

Webinars:
- Beyond Recall: Taking Competency Assessments to the Next Level
- Item Analysis for Beginners
- 10 Quick Tips to Improve your Tests and Exams
- Three Dimensions of Planning Reliable Tests and Exams (covers blueprints well)

Upcoming webinars

Introduction to Questionmark's Assessment Platform

◆ September 16, 2021 - 10:00 am to 11:00 am (EDT)

Learn the basics of authoring, delivering and reporting on surveys, quizzes, tests and exams. This introductory webinar explains and demonstrates key Questionmark features and functions.

Click to Register

Making Scores Meaningful: The Role of Standards

◆ September 23, 2021 - 11:00 am to 12:00 pm (EDT)

This session will explore some key concepts in understanding assessment practices. Using examples from the driving test, to educational qualifications, to job interviews, you will see the real-world applicability of these concepts, and see their relevance to one's own assessment decision making.

Click to Register

The Future of Learning: Is a Hybrid Model Here to Stay?

◆ October 13, 2021 - 11:00 am to 12:00 pm (EDT)

Presented by Ivan Babovic of Proctorio and John Kleeman of Questionmark, this session will look into the future of validation of higher education and learning by focusing on value to the learner.

Click to Register
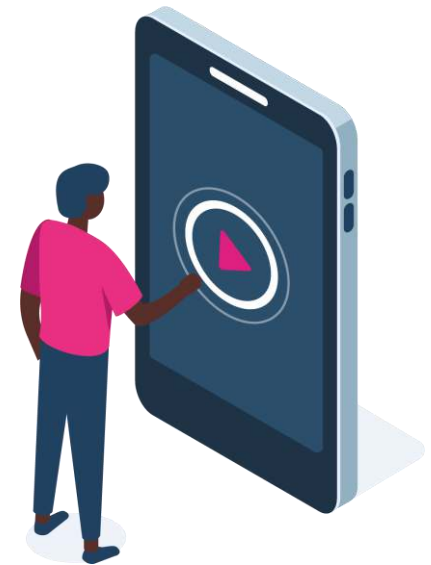
question mark

Thank you for attending!

We hope to see you at a future webinar

Keep up to date at www.questionmark.com/resources/blog