Questionmark

powered by **Learnosity**

Did They Learn It?
Can They Do It? Can You Prove It?

Getting Trustworthy, Defensible Results when Assessing
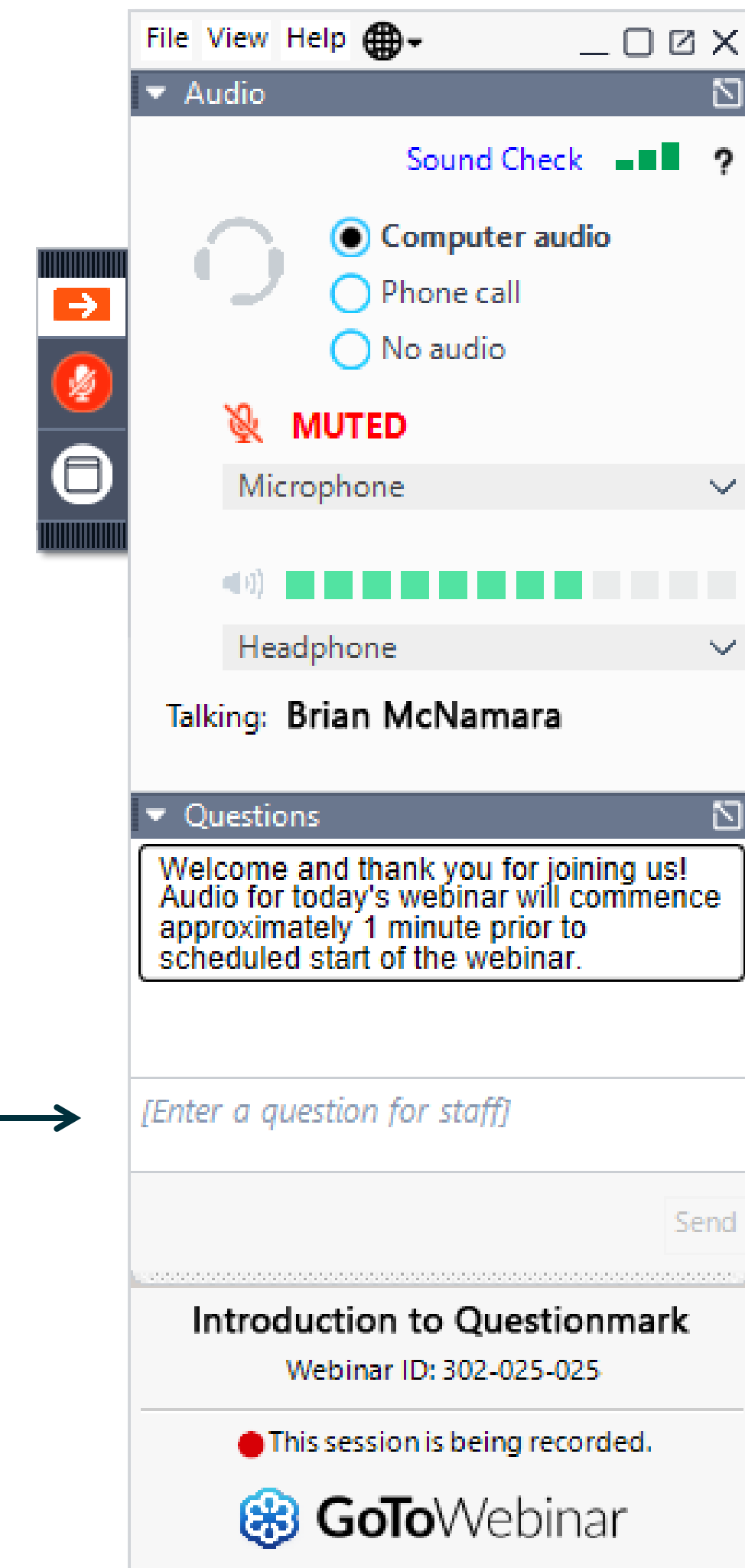Knowledge, Skills and Abilities

Before we get started...

Watch for an email after the webinar to:

- Download slides (PDF)
- View a recording

To ask questions, use the "Questions" feature

File **View** **Help** 🌐▾

▾ Audio

Sound Check ▪▪▪ **?**

◉ Computer audio
○ Phone call
○ No audio

🔇 **MUTED**

Microphone ▾

🔊 ▪▪▪▪▪▪▪▪

Headphone ▾

Talking: **Brian McNamara**

▾ Questions

Welcome and thank you for joining us!
Audio for today's webinar will commence
approximately 1 minute prior to
scheduled start of the webinar.

[Enter a question for staff]

Send

Introduction to Questionmark
Webinar ID: 302-025-025

🔴 This session is being recorded.

GoToWebinar

Questionmark
powered by Learnosity

Questionmark's Mission

Questionmark's mission is to provide the highest quality testing and assessment software and support services to enable individuals and organizations reach their goals.

What can we learn from assessments?

Learn

Do

Prove

Questionmark
powered by Learnosity
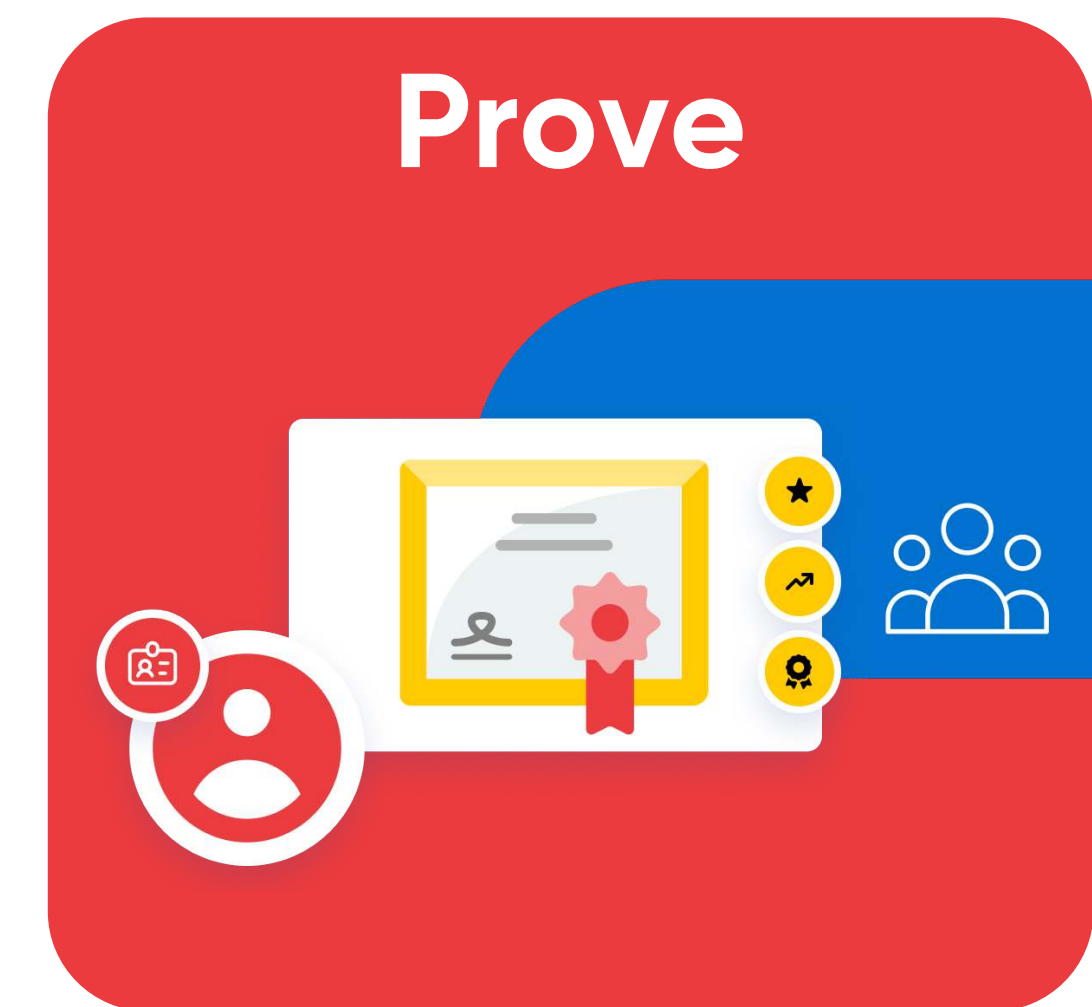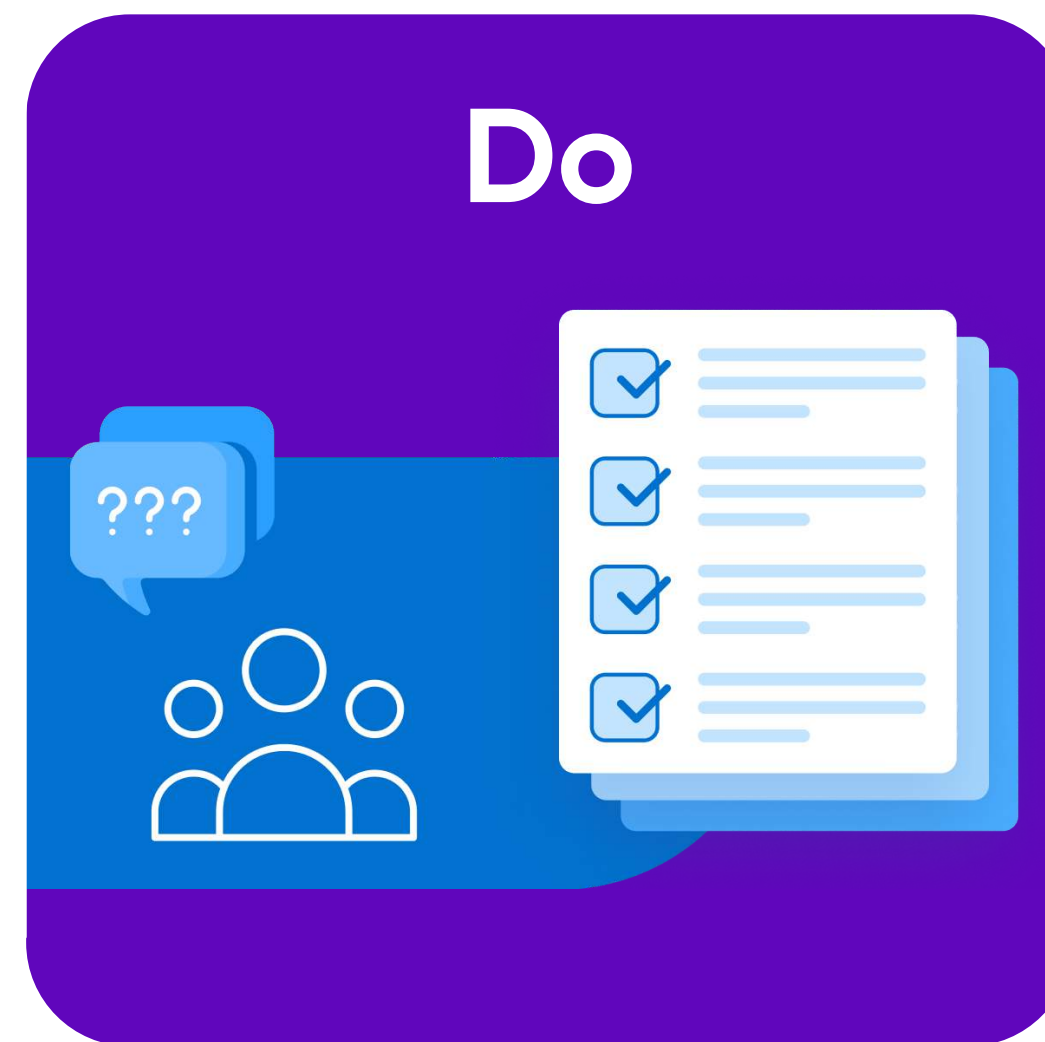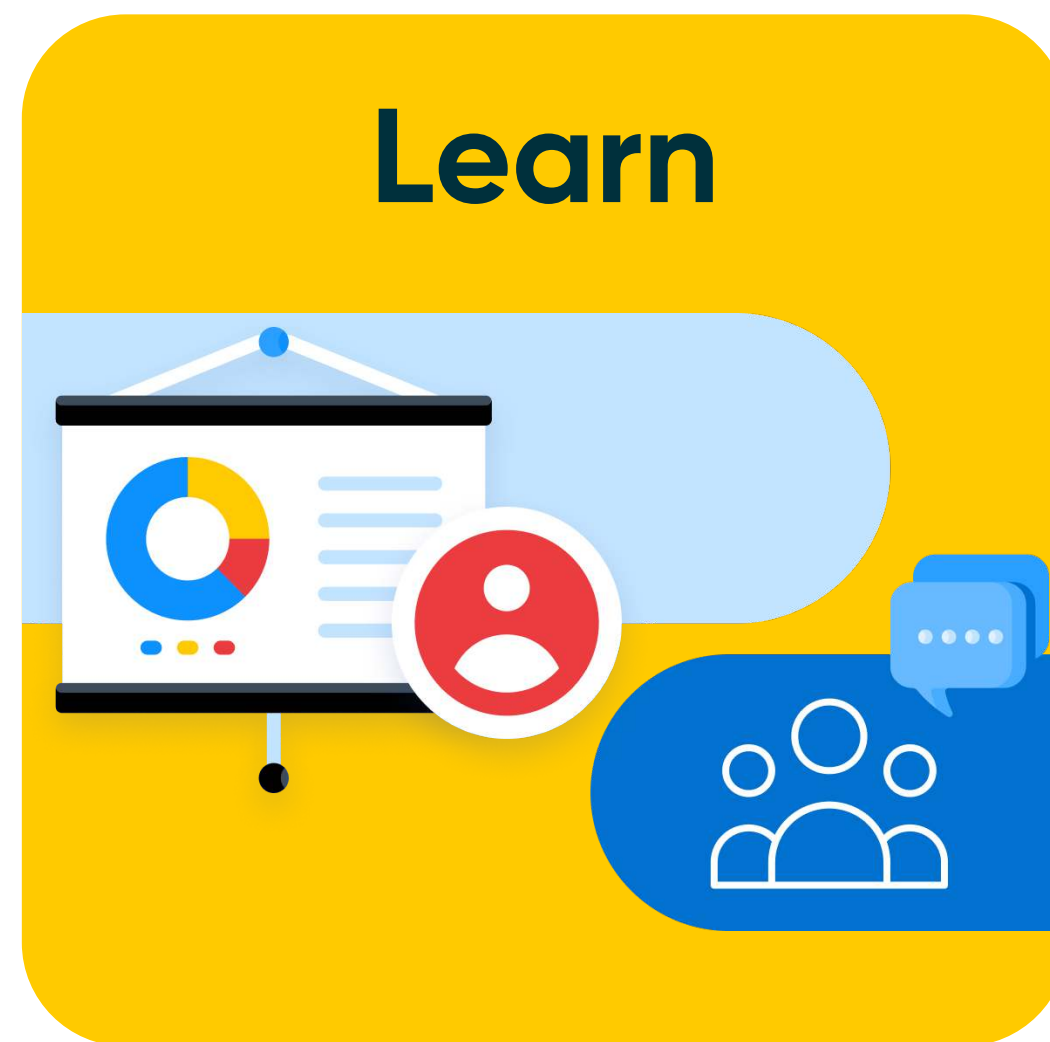
What we'll cover today

Strategic role
of assessments

Importance of
validity and reliability

Planning an
assessment

Measuring
"above knowledge"

Assessments and
evaluating learning

Questionmark
powered by Learnosity

Assessments provide strategic benefits

- Make informed, defensible personnel decisions

- Mitigate regulatory compliance risk

- Align workforce competencies with organizational objectives

- Target learning investments accurately

- Onboard new employees effectively

- Verify skills/knowledge vital organizational KPIs

- Enhance learning/knowledge retention

Questionmark
powered by Learnosity
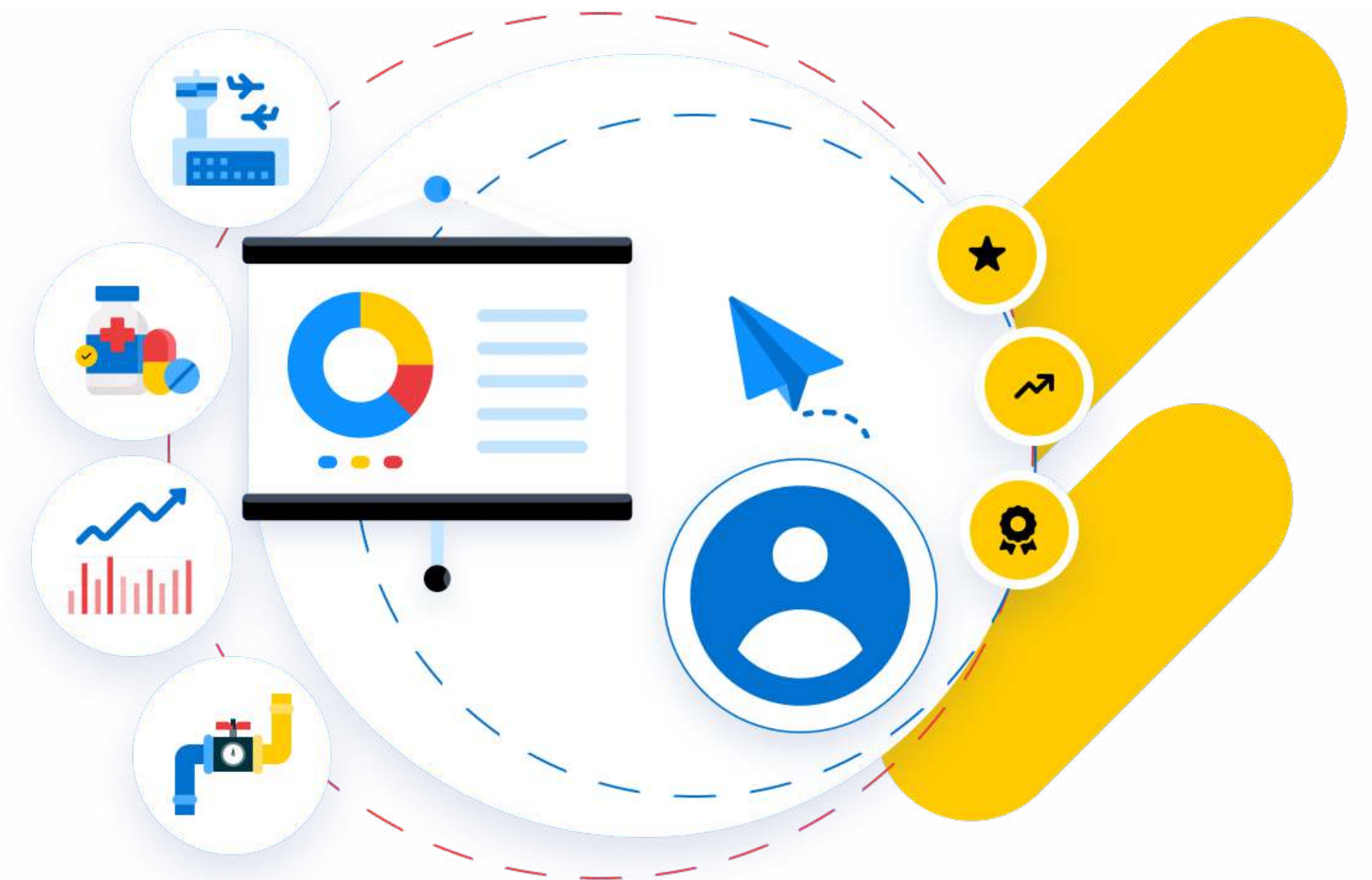
Stakes of assessments

- Life, limb and livelihood
- Reputation

Banks fined £2bn over forex scandal

Auto Manufacturer to Spend $14.7 billion to Settle Emissions Case

Bank Fined US$185 Million over unauthorized accounts

Roofer's first-day injury costs Florida contractor $152K OSHA fine

Universities fined millions for how labs how their labs managed chemical waste

Questionmark
powered by Learnosity

Workplace assessment: Stakes of pre/post-hire testing

An HR decision made based on the outcome

The decision has an impact on a person's employment conditions

Uniform Guidelines for Employee Selection Procedures (US EEOC, 1978): "Employment decisions include but are not limited to:
- Hiring
- Promotion
- Demotion
- Membership (for example in labor organization)
- Referral
- Retention
- Licensing and Certification

"Selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above"

Questionmark
powered by Learnosity

Three foundations for good tests

Validity

- Degree to which evidence and theory support the interpretation of test scores for proposed uses of tests

Reliability (or precision)

- Consistency of scores across instances of the testing procedure
- Reduced measurement error

Fairness

- Fair and equitable treatment of all individuals in the intended population of test-takers
- Does not advantage or disadvantage individuals because of characteristics irrelevant to the construct being measured
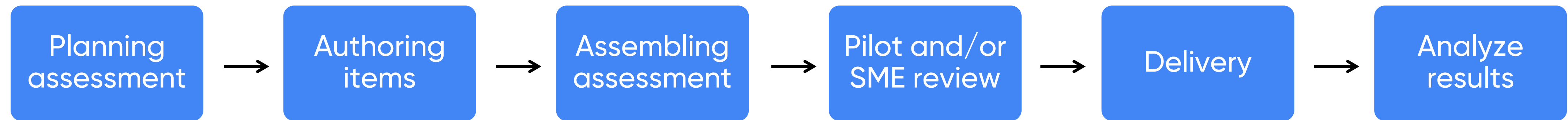
Based on "The Standards for Educational and Psychological Testing" AERA/APA/NCME

How do we get there?

Typical assessment development process

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

Questionmark
powered by Learnosity

Document the purpose of your test

Why?

- Why are you delivering the test?

- Whether norm referenced or criterion referenced

What?

- What construct or domain is being measured?

Who?

- Who is taking the test?
- What are their language and computer skills?
- What diversity/fairness issues are important?

How?

- What action if someone passes?
- What action if someone fails?
- How else will you use the scores?

Questionmark
powered by Learnosity

Determine content of the test based on the purpose

End of course test	• Course content and objectives
Placement tests	• Entry-level knowledge and skills
Certification tests, employment tests	• Job Task Analysis, job requirements

- Develop a test blueprint (AKA "test content outline")

- Covers what is included and what is excluded

- Often a series of objectives and a weighting

- May include key knowledge or skills areas to include

Questionmark
powered by Learnosity

12

Job Task Analysis

Identify tasks and behaviors

Identify conditions and environment

Identify Knowledge, skills and abilities required

Methods

- Panel of experts to describe the job
- Panel of stakeholders to define expectations
- Interview experts and stakeholders
  - What is done?
  - Why is it done?
  - Why is it important?
- Survey experts and stakeholders to identify trends or patterns
- Review related literature and documentation

Questionmark
powered by Learnosity

Job Task Analysis (JTA) Surveys for content planning, validity

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

Survey SMEs about key tasks:

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

What is your role in the organization? [⌄]

3 of 3
Answer questions about nursing.

	Applicability			Difficulty				
	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	○	○	○	○	○	○	○	○

Questionmark
powered by Learnosity

Simple example of assessment blueprint

Electrical Skills Test – Blueprint	% of test	# items	Easy	Medium	Hard
Topic 1: AC Motors	45%	**34**			
Subtopic 1.a		21	8	8	5
Subtopic 1.b		8	2	5	1
Subtopic 1.c		5	2	2	1
Topic 2: High-voltage switchgear	25%	**19**			
Subtopic 2.a		12	3	5	4
Subtopic 2.b		7	2	4	1
Topic 3: Synchronous Motors	30%	**23**			
Subtopic 3.a		11	2	7	2
Subtopic 3.b		12	4	5	3

Approaches to developing test blueprints

Process-Based Blueprints

- Define processes to be assessed
  - Cognitive processes
  - Processes related to job performance
- 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. . ."

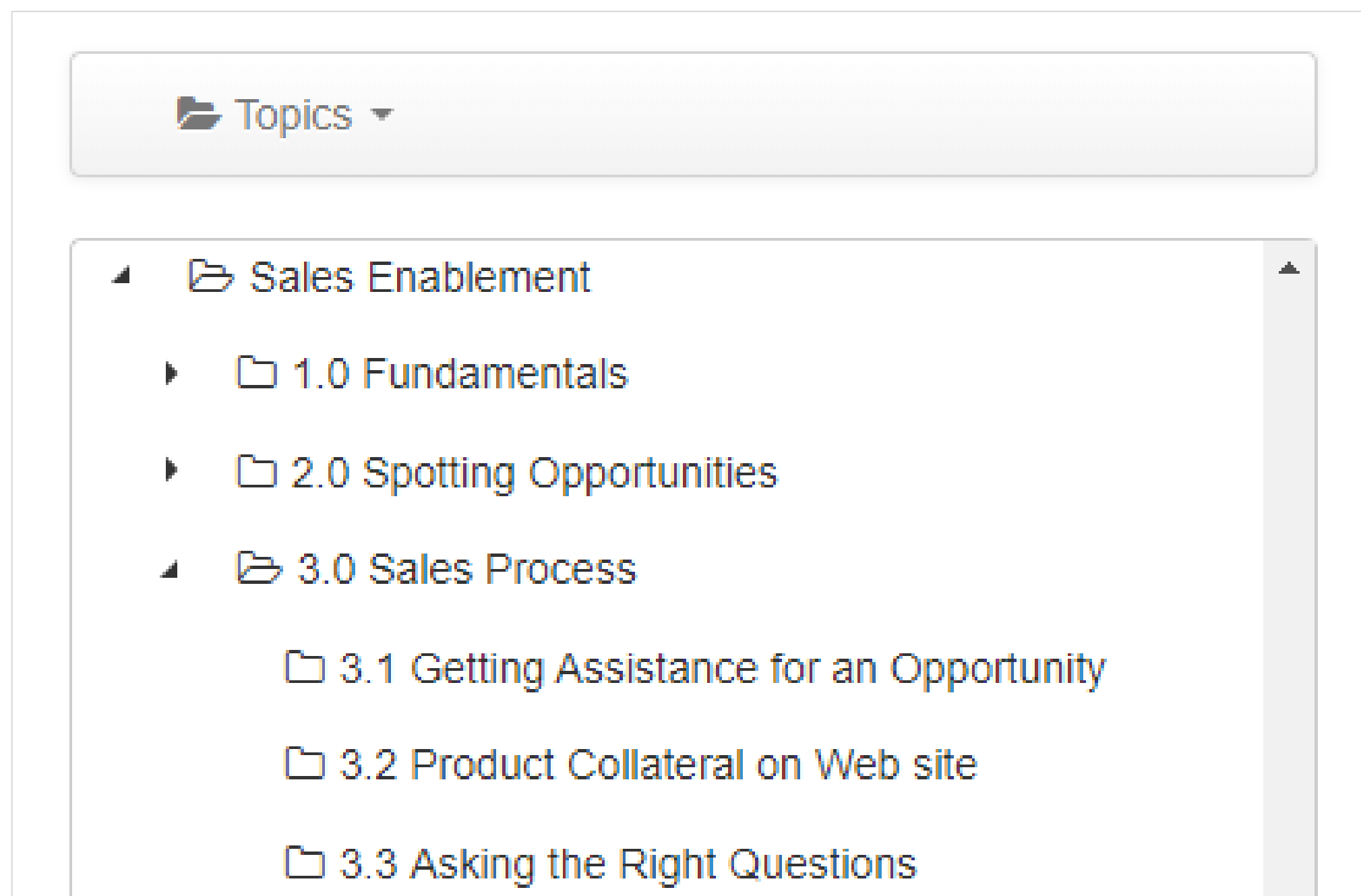Provides guidance on types of behaviors items should be designed to elicit

Content-Based Blueprints

- Define content to be assessed, such as:
  - Curricular or training content
  - Problems often encountered on the job
- Based on nouns that identify topics (Raymond & Neustel, 2006)
  - "Algebra"
  - "Client Support"

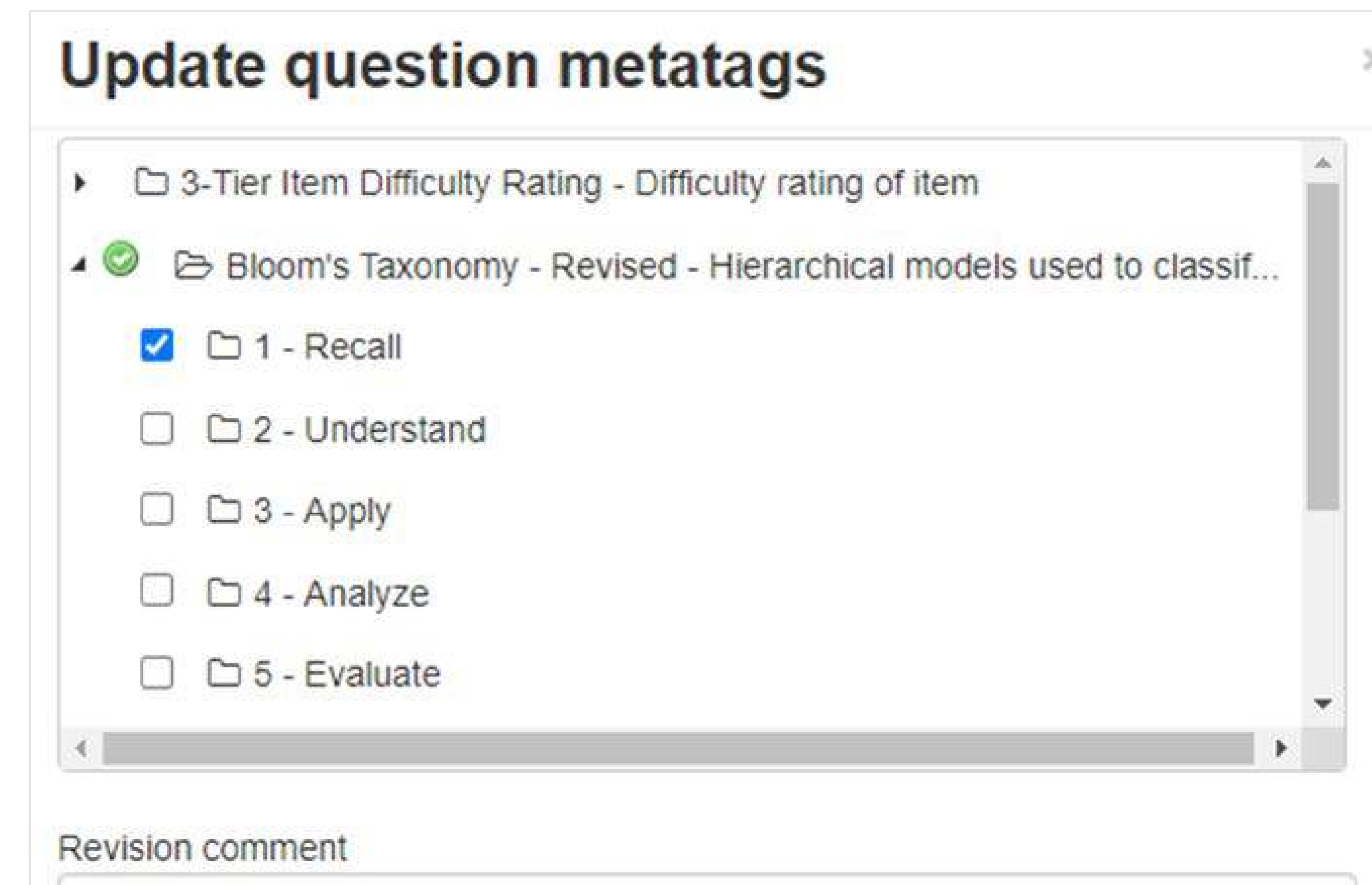Easy to communicate content of the assessment, easily mapped to curriculum

Questionmark
powered by Learnosity

16

Relevant Questionmark features

Questions Organized by Topic/Subtopic

Meta tags for additional categorization

Content and process together

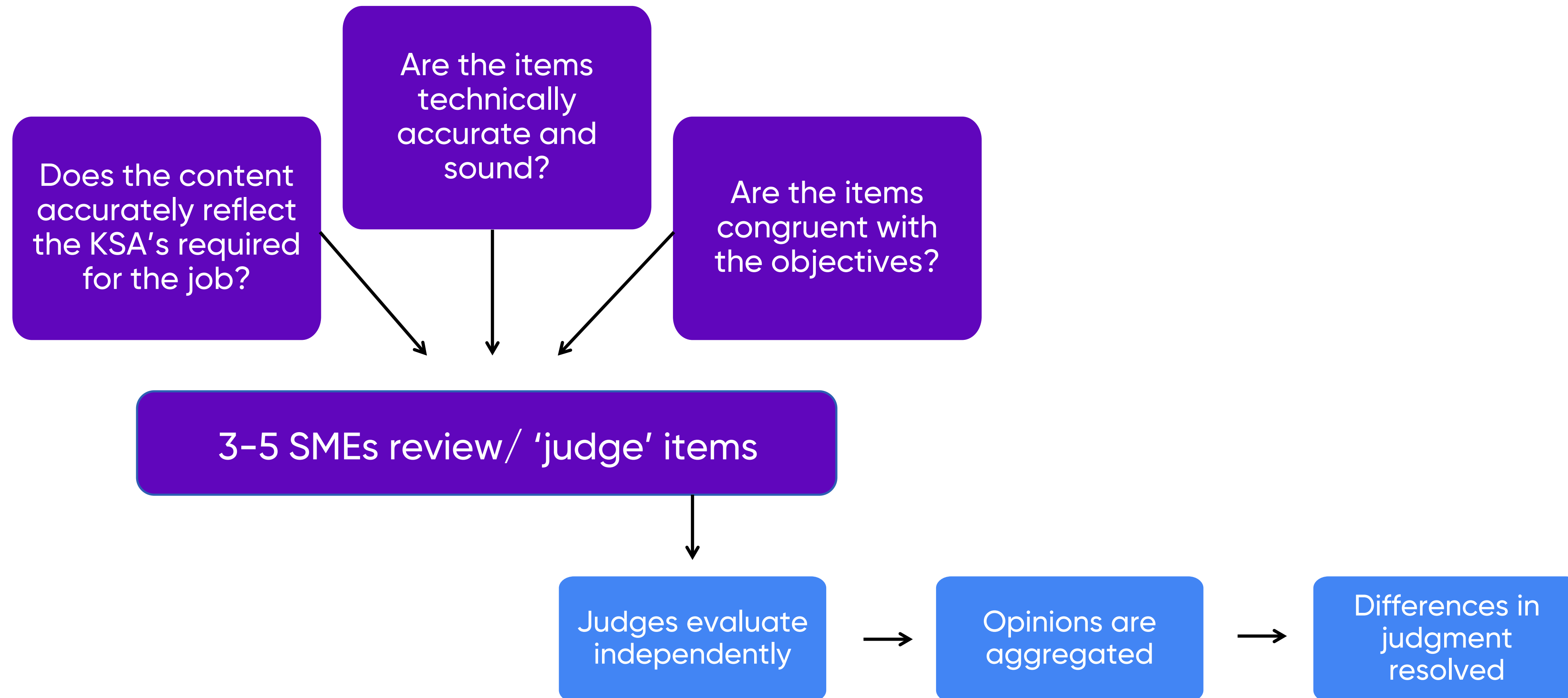| Content	Weight	# Items	Cognitive Level					
			Remember	Understand	Apply	Analyze	Evaluate	Create
Ingredients	50%	50	to list three ingredients in a PB&J	to explain the flavor and texture profiles of ingredients in a PB&J	to find the required ingredients at a grocery store	to differentiate between unlabeled jars of crunchy PB and creamy PB	to contrast the ingredients used in a PB&J and a Ve[...] 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
			4	30				

Cognitive Levels (Questionmark meta tags)

Content (Questionmark Topics)

Weight on assessment

Processes (i.e. Learning Objectives)

Questionmark
powered by Learnosity

Content validity review

Does the content accurately reflect the KSA's required for the job?

Are the items technically accurate and sound?

Are the items congruent with the objectives?

3-5 SMEs review/ 'judge' items

Judges evaluate independently → Opinions are aggregated → Differences in judgment resolved

Questionmark
powered by Learnosity

Assessment development: A process and a cycle

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

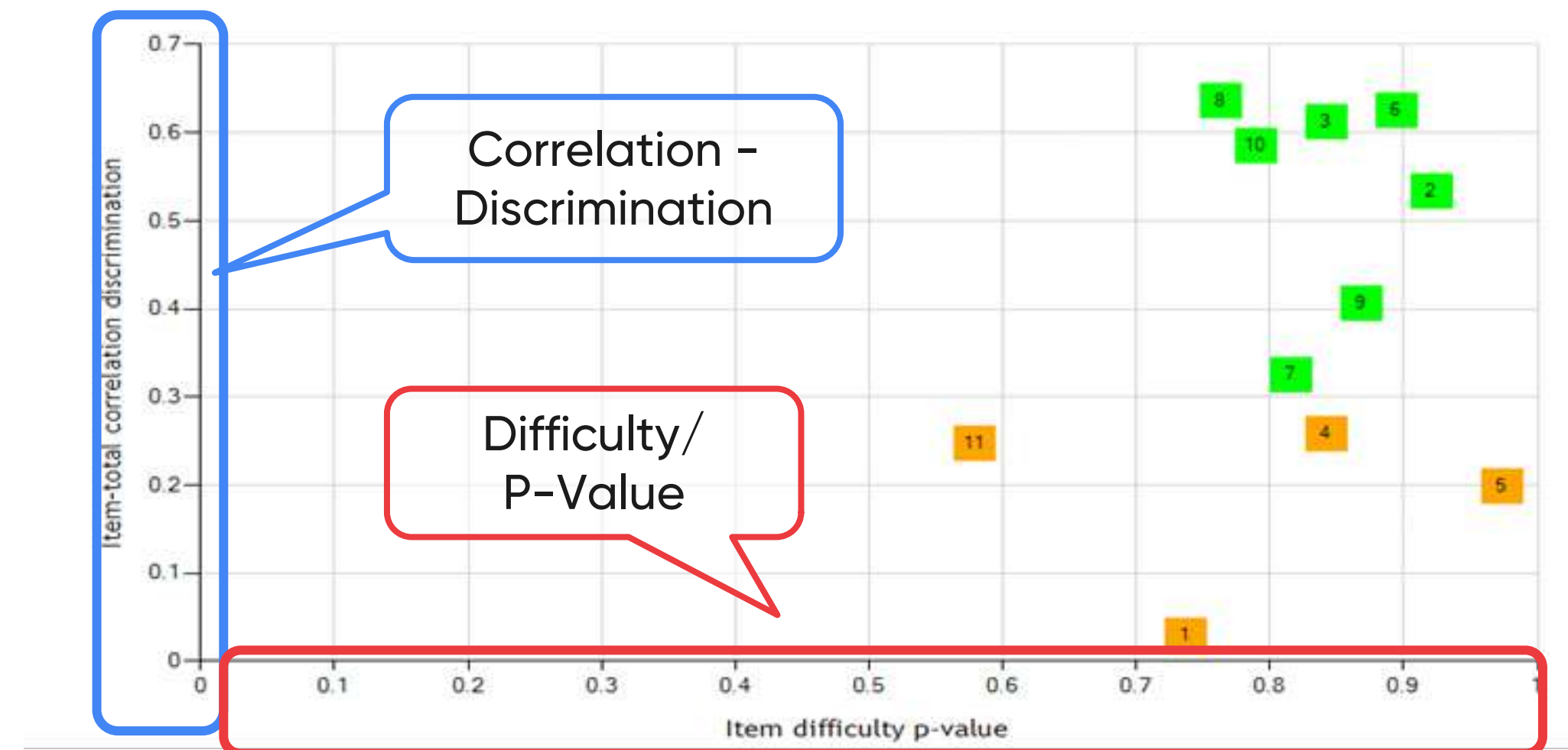**Questionmark**
powered by Learnosity
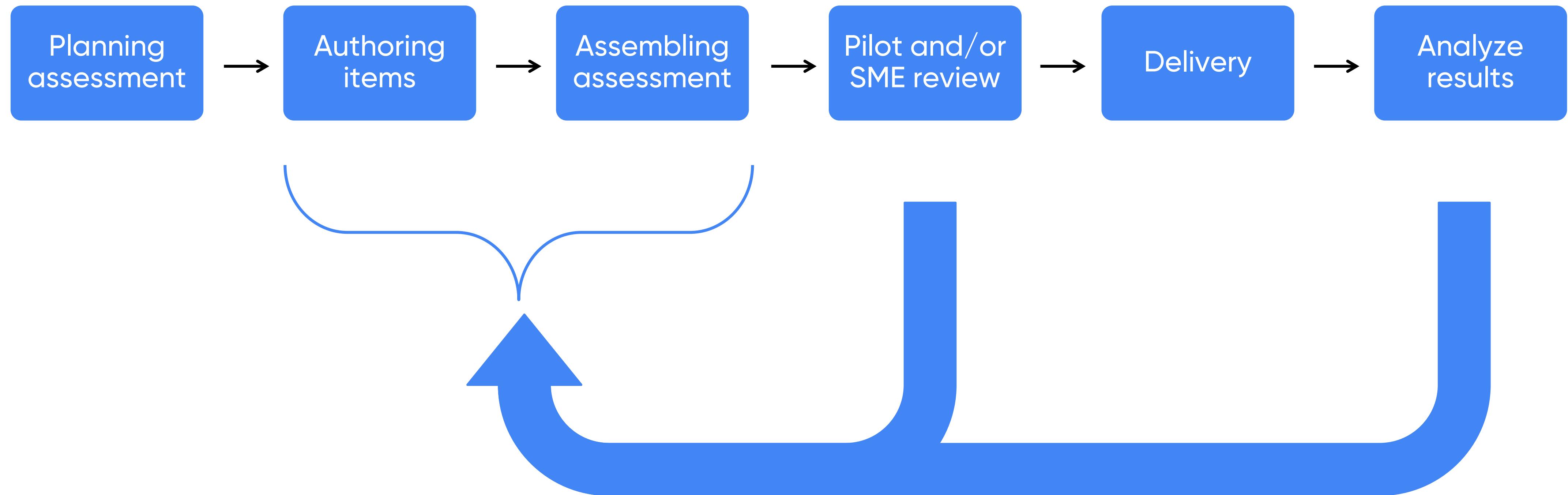
Analyze item response statistics

Difficulty index ("P value")

- ~% answered correctly
- Expressed as from 0.0 to 1.0
  - Too high: Too easy
  - Too low: Too difficult? Confusing/ambiguous? Miscoded?
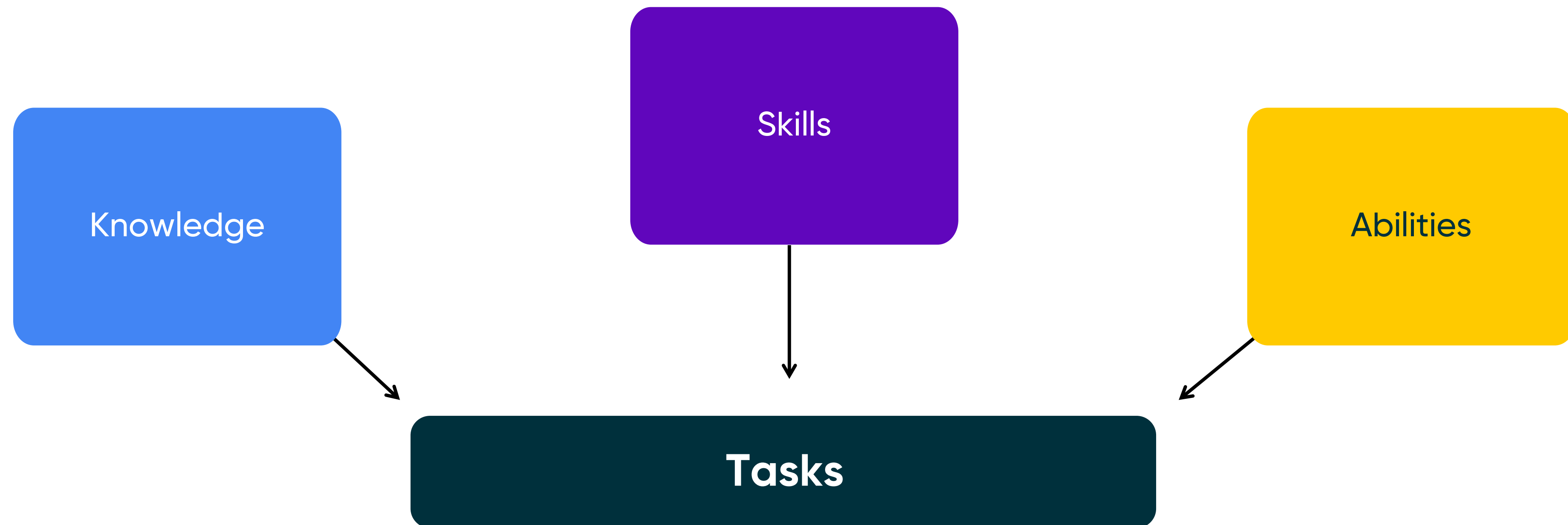
Correlation discrimination

- Correlation between **item** and **assessment scores**
- How effectively the question differentiates between participants of different knowledge/skill levels
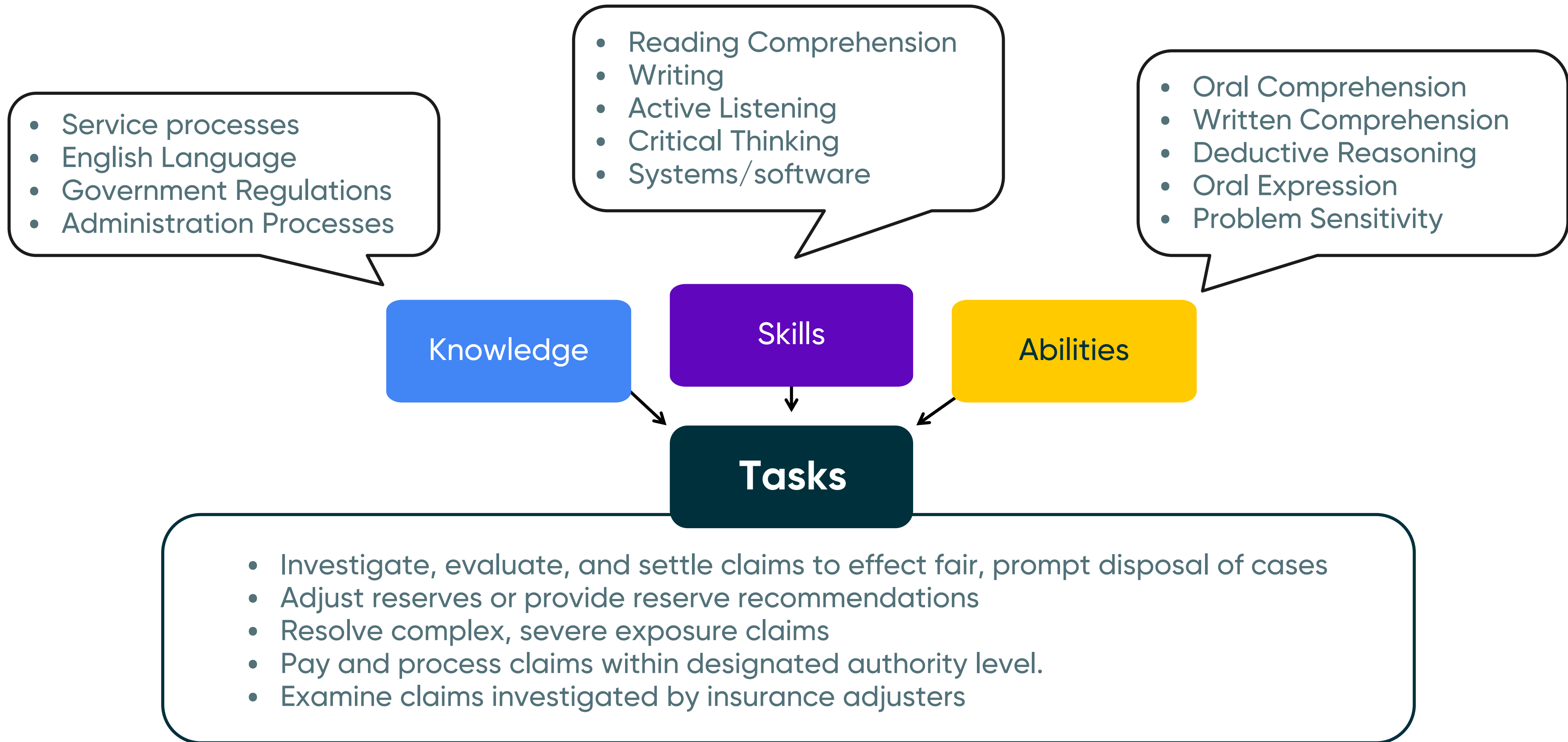

Question description	Revision	Topic	Item difficulty p-value	Item–total correlation discrimination
A multiple choice question must have at least four choices.	2	Item Writing Best Practices	0.737	0.03
In test questions, it is good practice to inlcude implausible, nonsense, or humorous, incorrect choices.	2	Item Writing Best Practices	0.921	0.534
Phrasing a question as, Which of the following is not ... is considered best practice.	3	Item Writing Best Practices	0.842	0.613

Questionmark
powered by Learnosity

Measuring Above Knowledge

Competencies and performing tasks

Example: Insurance Claims Examiner

- Service processes
- English Language
- Government Regulations
- Administration Processes

- Reading Comprehension
- Writing
- Active Listening
- Critical Thinking
- Systems/software

- Oral Comprehension
- Written Comprehension
- Deductive Reasoning
- Oral Expression
- Problem Sensitivity

Knowledge **Skills** **Abilities**

Tasks

- Investigate, evaluate, and settle claims to effect fair, prompt disposal of cases
- Adjust reserves or provide reserve recommendations
- Resolve complex, severe exposure claims
- Pay and process claims within designated authority level.
- Examine claims investigated by insurance adjusters

Questionmark
powered by *Learnosity*

Competencies measured

Knowledge (a cognitive Skill)
- Memory is a cognitive skills that provides ability to recall knowledge:
  - "Knowledge How"
  - "Knowledge To"

Above Knowledge Skills
- Cognitive skills above recalling knowledge
- (Comprehension , Application, Analysis, Synthesis and Evaluation)

Observable Skills
- Motor Skills
  i.e. physical movement/coordination
- Inter-personal skills

Attitudes and Behaviors
- Motivation
- Ability to achieve goals
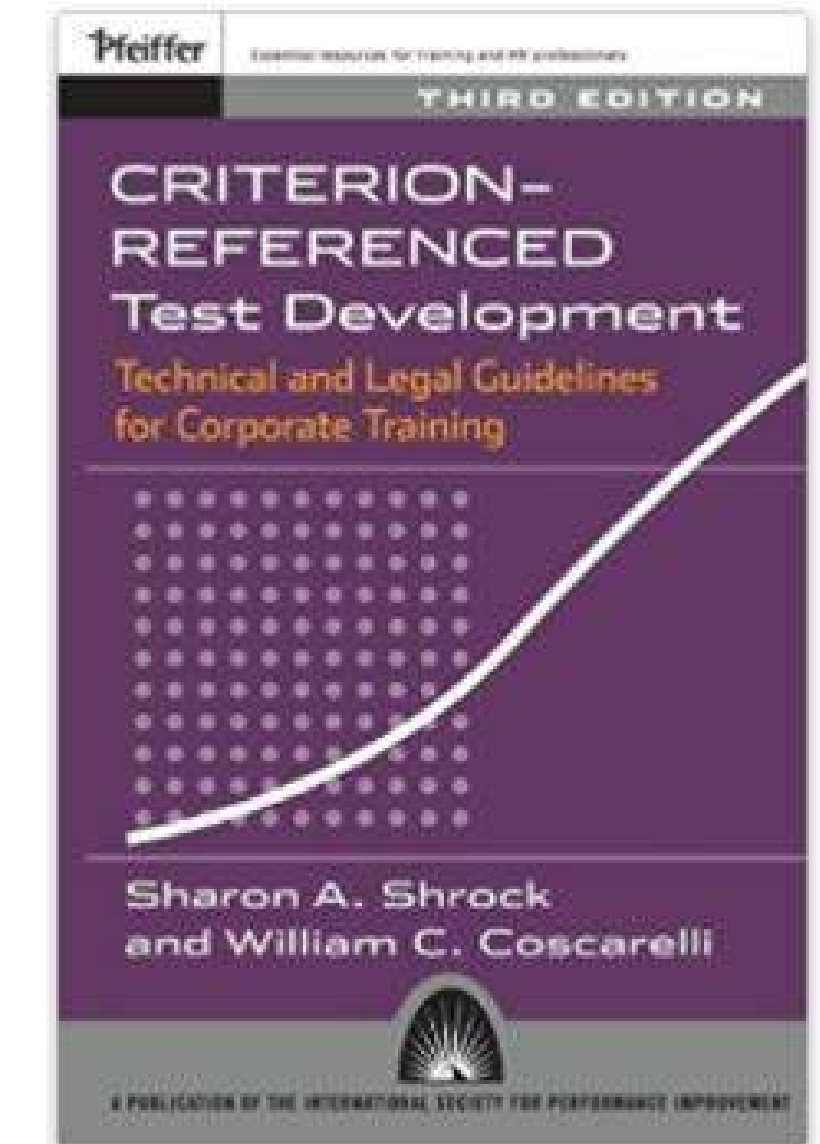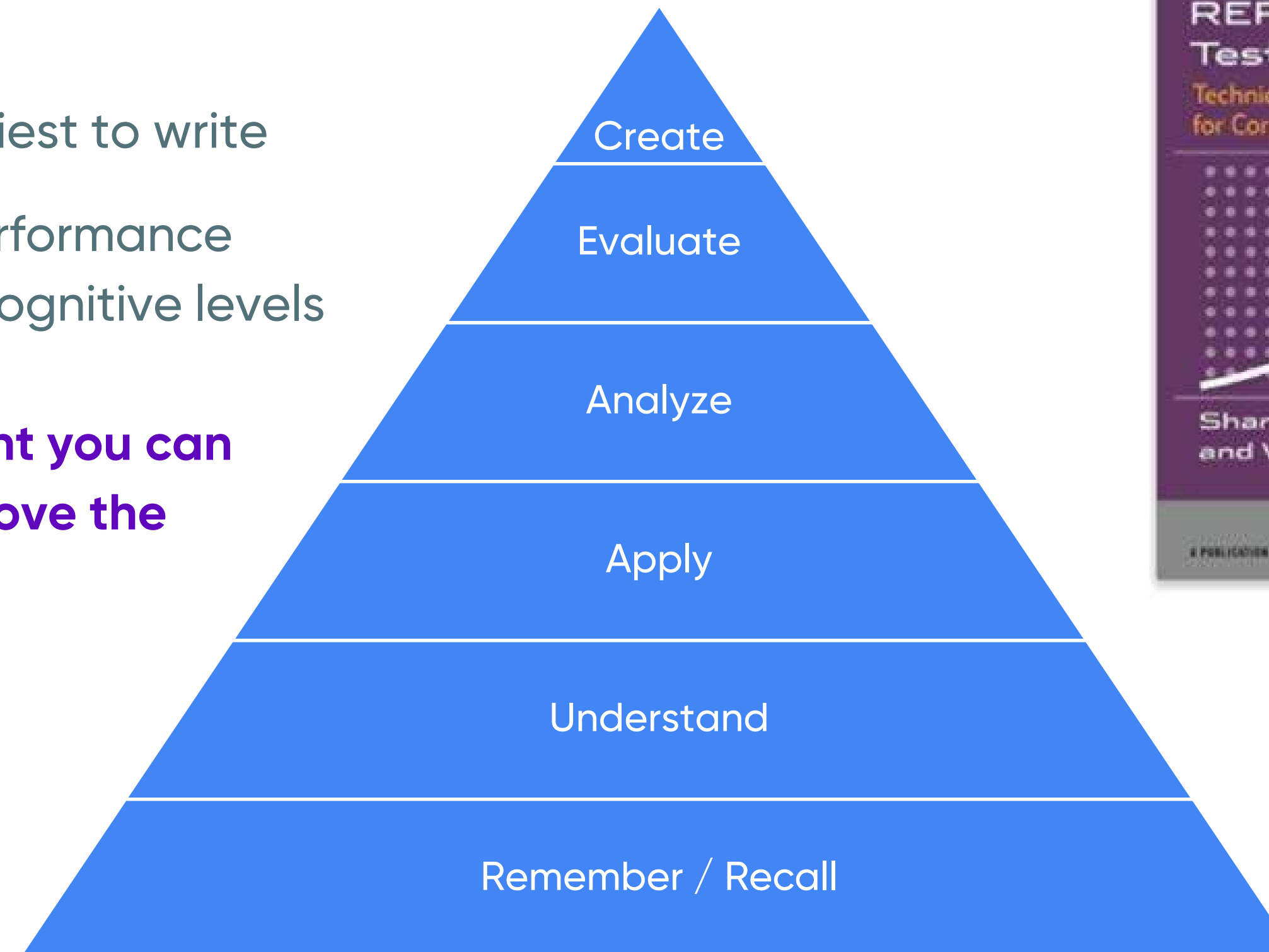- Team fit

Ability to Do/Perform "Competency"

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 above the memorization level.

Create
Evaluate
Analyze
Apply
Understand
Remember / Recall

CRITERION-REFERENCED Test Development
Technical and Legal Guidelines for Corporate Training

Sharon A. Shrock and William C. Coscarelli

Some approaches to assessing higher level skills

Writing conventional questions to assess beyond recall

- Use multiple choice, matching, ranking and other question types
- Techniques eliciting higher level cognition required to get the right answer
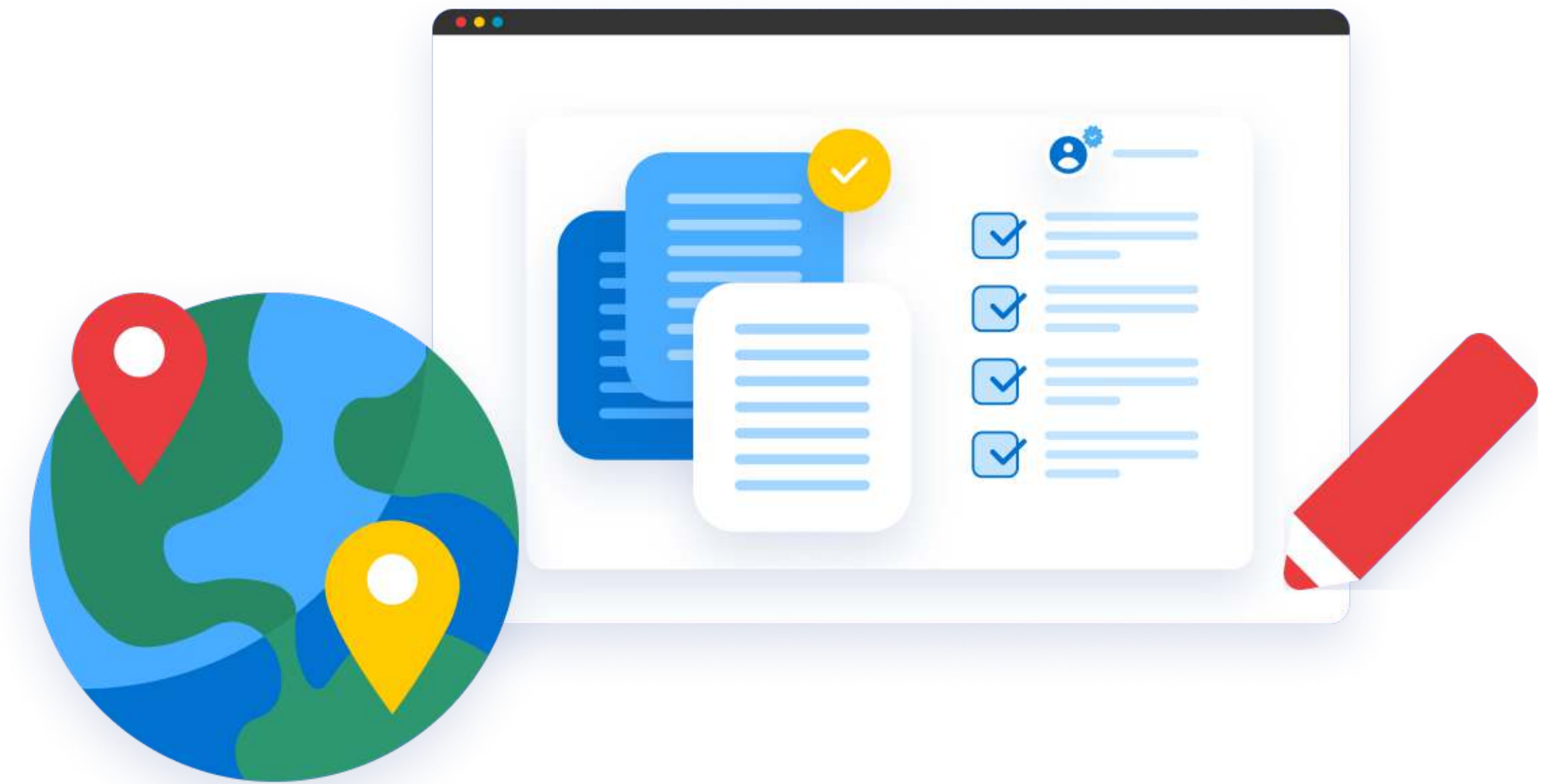
Observational Assessment / Performance Testing

- An observer watches someone performing a practical task
- Answers questions about performance on a mobile device

Situational Judgement Assessments

- Frame questions based on a dilemma that could be encountered in workplace
- Evaluate judgement of what course of action is taken

Questionmark
powered by Learnosity

Writing conventional questions to assess beyond recall

- Questions that require participant to consider multiple facts

- 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

Questionmark
powered by Learnosity

Observational / Performance Assessment

Observer "rates" the participant

Rate the participant's execution of operator tasks outlined in section 2.3.1.

- o Performed without prompting
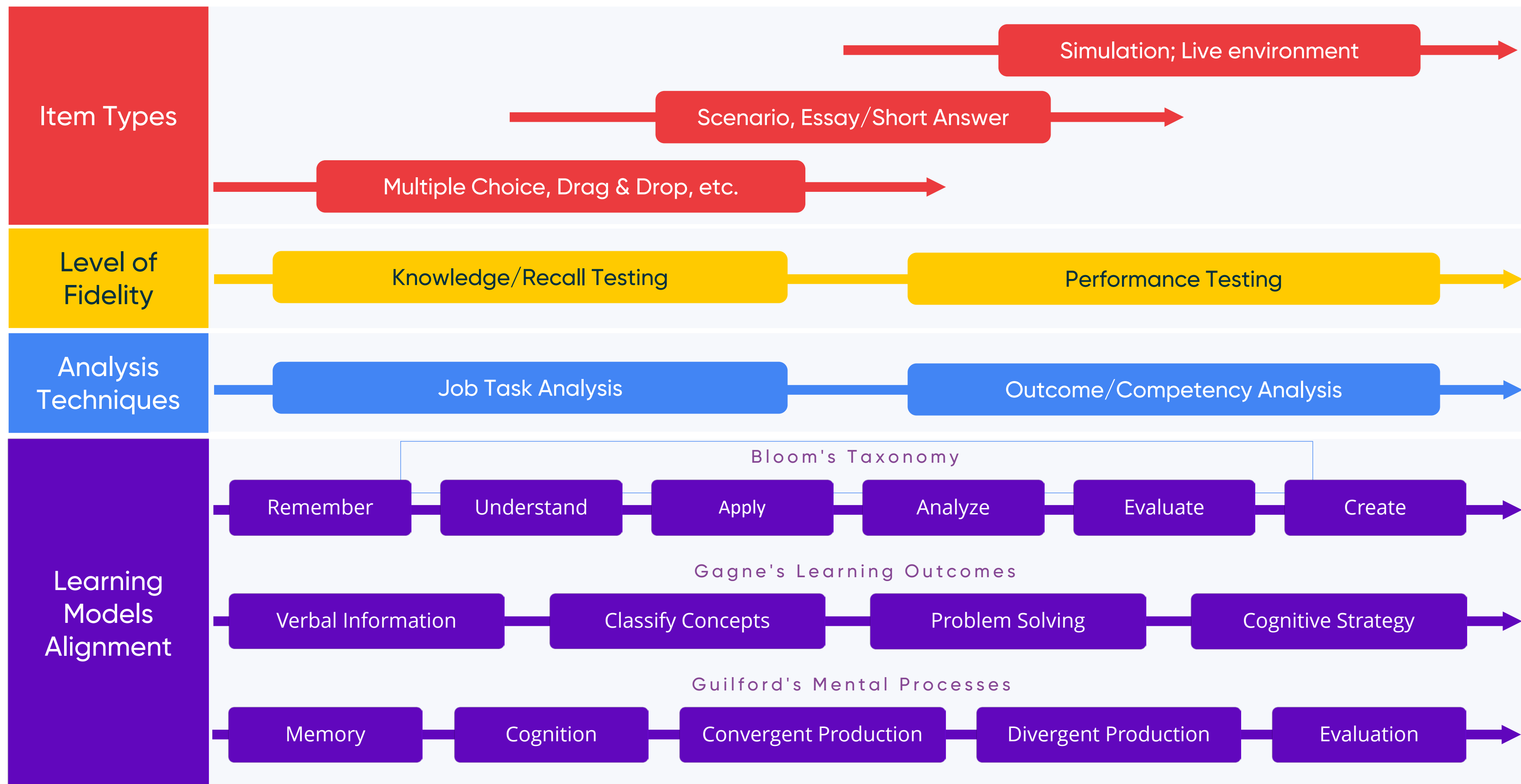- o Performed with prompting
- o Unable to perform

Item solicit ratings

- Performance of task(s)
- Demonstration of appropriate knowledge, skills, abilities, behaviors

Getting objective ratings

- Enforces consistent rating scales & scoring rubrics
- Streamlines and centralizes collection of input from raters

Equipment Operator Observational Assess...

1 of 3

Rate the participant's execution of operator tasks outlined in section 2.3.1.

- Performed without prompting
- Performed with prompting
- Unable to perform

2 of 3

Rate the participant's execution of operator tasks outlined in section 5.2

- Performed without prompting
- Performed with prompting
- Unable to perform

3 of 3

Rate the participant's execution of operator tasks outlined in section 1.5

- Performed without prompting
- Performed with prompting

Questionmark
powered by Learnosity

Performance Testing & Development Taxonomy

Item Types

- Simulation; Live environment
- Scenario, Essay/Short Answer
- Multiple Choice, Drag & Drop, etc.

Level of Fidelity

- Knowledge/Recall Testing
- Performance Testing

Analysis Techniques

- Job Task Analysis
- Outcome/Competency Analysis

Learning Models Alignment

Bloom's Taxonomy

Remember	Understand	Apply	Analyze	Evaluate	Create

Gagne's Learning Outcomes

Verbal Information	Classify Concepts	Problem Solving	Cognitive Strategy

Guilford's Mental Processes

Memory	Cognition	Convergent Production	Divergent Production	Evaluation

Investing in sound assessment development processes and high-quality items can offer dividends

Case study

Benefits of "Testing Out" of training

Management Report Example	Course Only
Number of people	1,000
Time required for ALL to complete training if no testing (2 hours each)	2,000 hours
Time for ALL to take test (20 mins / test)	
25% pass test and skip training; Remaining 75% do required training	
Total potential billable hours	2,000
Average cost per billable hour	$100
Total costs in billable hours	**$200,000**

Questionmark
powered by Learnosity

Benefits of "Testing Out" of training

Management Report Example	Course Only	Test & Course
Number of people	1,000	1,000
Time required for ALL to complete training if no testing (2 hours each)	2,000 hours	
Time for ALL to take test (20 mins / test)		333 hours
25% pass test and skip training; Remaining 75% do required training		1,500 hours
Total potential billable hours	2,000	1,833
Average cost per billable hour	$100	$100
Total costs in billable hours	**$200,000**	**$183,300**

8% reduction
in billable hours
lost saving $16,700

Questionmark
powered by Learnosity

Course A
(25 Questions)

→

Course B
(28 Questions)

→

Course C
(21 Questions)

Getting valid results in less time
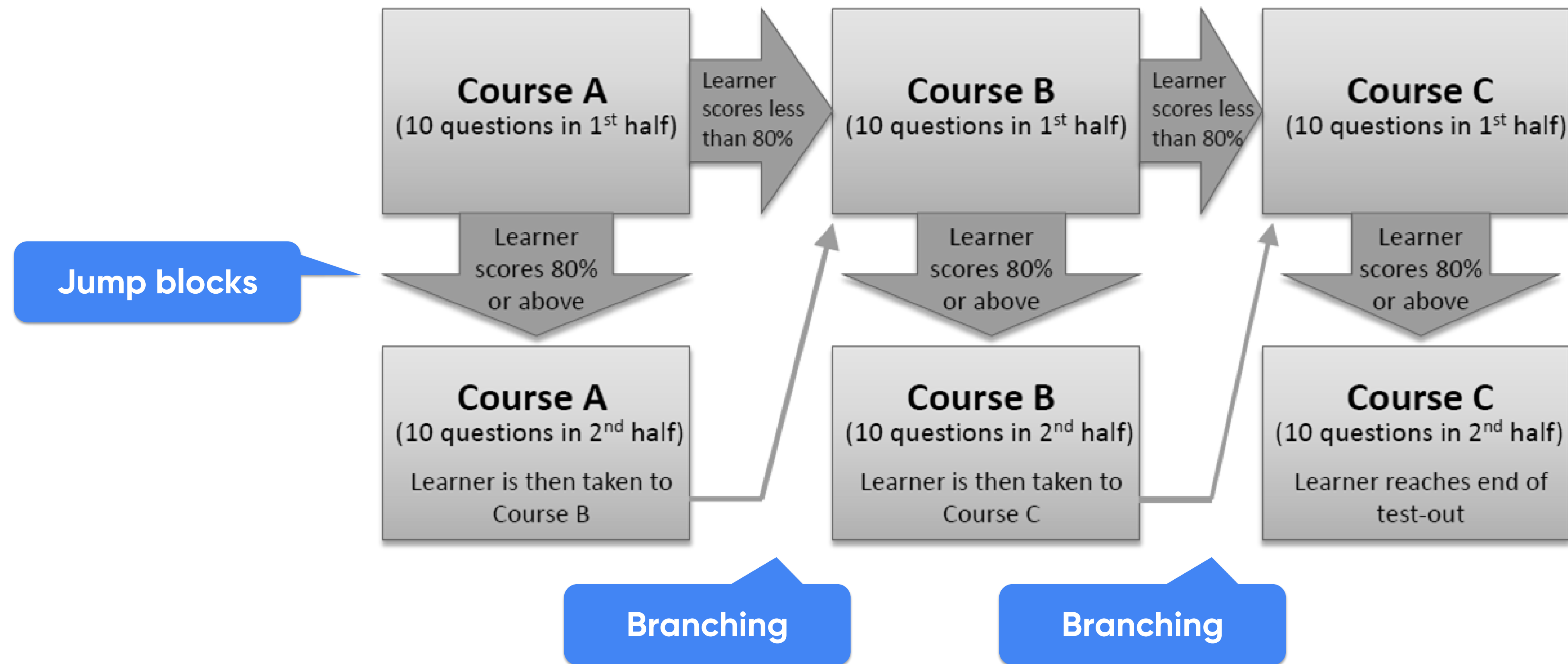
Lessons learned

- **Item analysis to:**
  - Quickly spot/flag and weed out problematic test questions
  - Demonstrates reliability of item/assessment results
  - Potential for shorter tests/exams

- **Branching logic within longer tests/exams**

- **Up-front investment in higher quality items/assessments reduce costs**
  - "Test out" of required training
  - Tests are shorter / less time away (means increase in billable hours)

Questionmark
powered by Learnosity

Pop quiz!

Which of the following is the most appropriate passing score for a knowledge test?

A. 70%

B. 80%

C. 90%

D. 100%

E. It depends on how difficult the questions are

Questionmark
powered by Learnosity

Setting Defensible Pass/Fail Scores

Setting Defensible Cut Scores

Risky practice:

- Guess
- Roll dice
- Pick a number out of a hat

Good Practice:

- Set pass/cut score to reflect minimally acceptable competence
- Passing test demonstrates competence
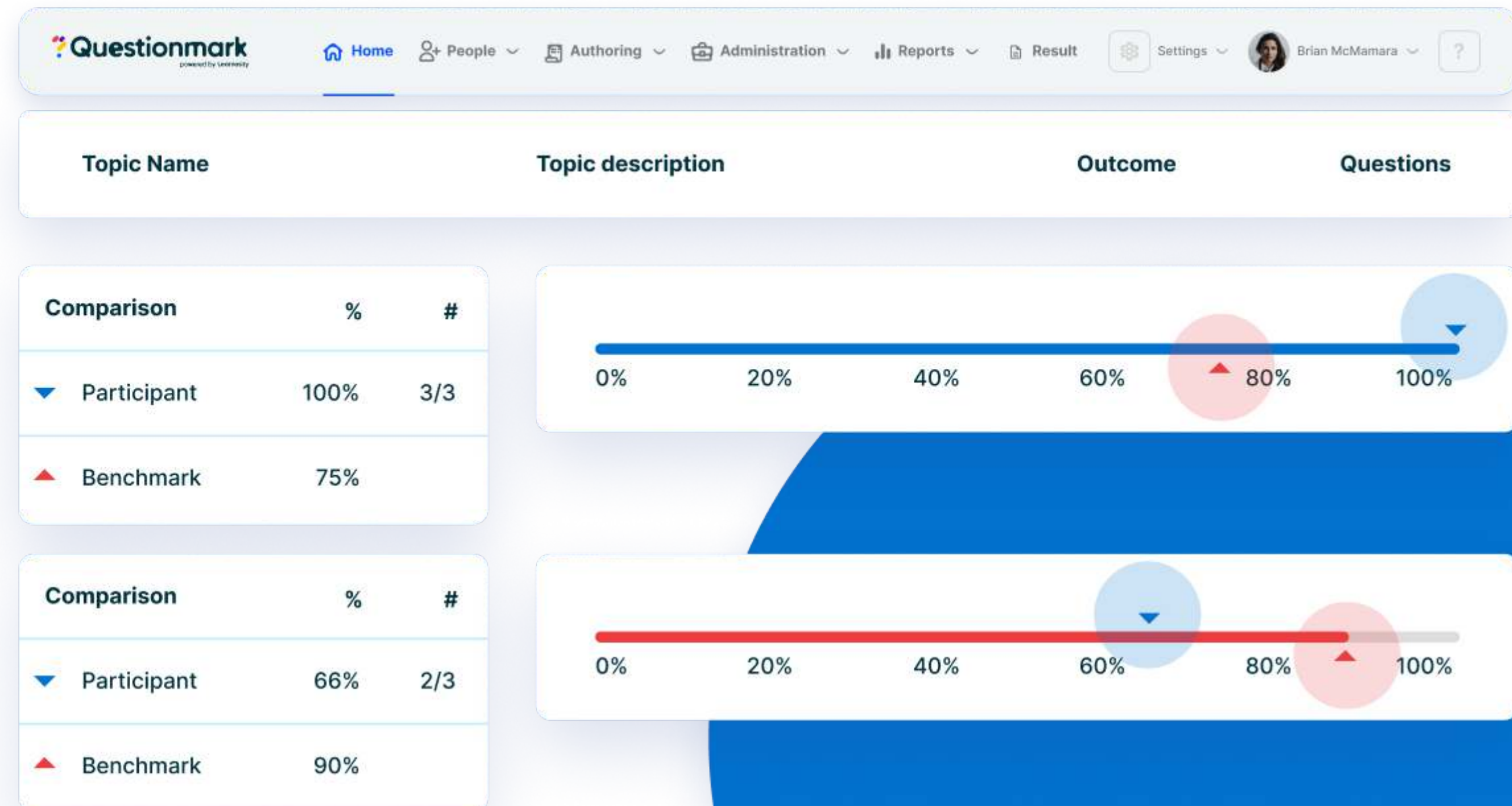
One route is the Angoff Method

- Based on this question:
  - What is % chance a marginal test-taker will get question right?
- How it works
  - Poll SMEs
  - Consider marginal test-takers and probability of getting specific questions right (0-100%)
  - Average out the chances to work out the cut score

Why use this method?

- Defensible
- Easy to use and implement
- Widely accepted

Questionmark
powered by Learnosity

Angoff Method example

What is the % chance that a borderline test-taker will get question right?

	SME A	SME B	SME C	Total
Q1	75%	75%	75%	75.00%
Q2	70%	80%	80%	76.67%
Q3	65%	75%	70%	70.00%
Q4	60%	85%	90%	78.33%
Q5	80%	80%	85%	81.67%
Q6	80%	80%	80%	80.00%
Q7	75%	80%	75%	76.67%
Q8	65%	90%	65%	73.33%
Q.....	75%	80%	75%	76.67%
Q50	65%	85%	65%	71.67%
Totals	71%	81%	76%	76%

Assessments and Evaluating Learning

The Kirkpatrick Model

Level 1

Reaction

Level 2

Learning

Level 3

Behavior

Level 4

Results

Questionmark
powered by Learnosity

The Learning-Transfer Evaluation Model

- Evolving model from Will Thalheimer
- Combines concepts from Blooms, Kirkpatrick and others

| Attendance | Activity | Learner Perceptions | Knowledge | Decision Making Competence | Task Competence | Transfer | Effects of Transfer |
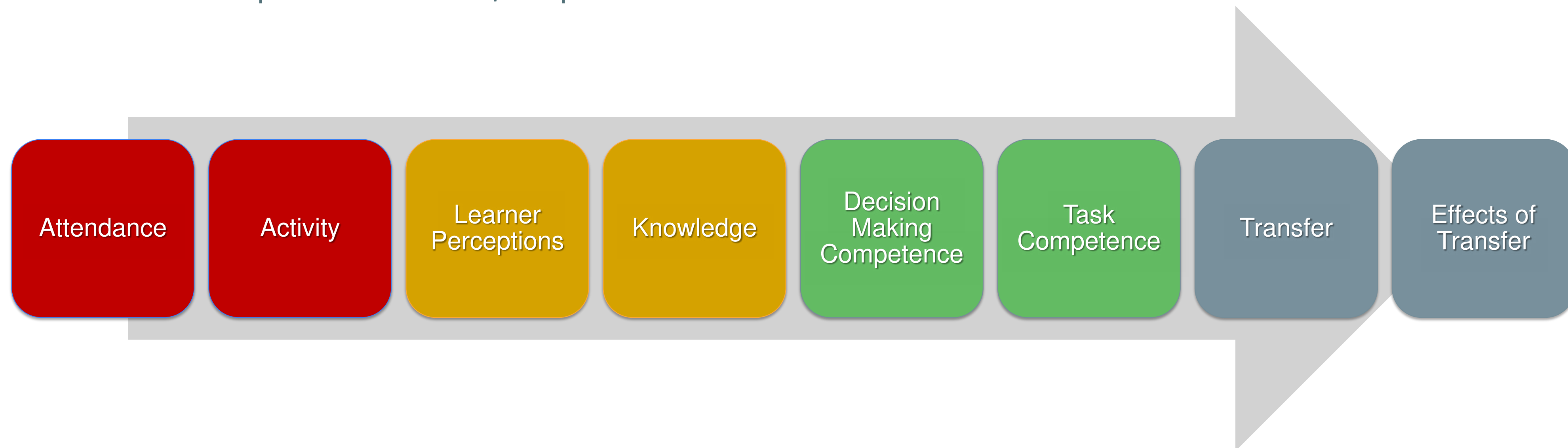
The Learning-Transfer Evaluation Model (LTEM) (version 12 / 17-May-2018)
https://www.worklearning.com/2018/02/14/the-learning-transfer-evaluation-model-ltem/

Questionmark
powered by Learnosity

Questions?

Useful resources

Useful resources
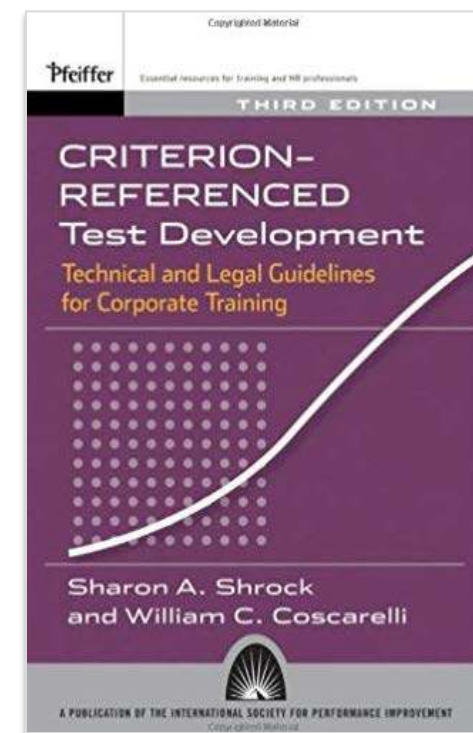
Criterion-Referenced Test Development: Technical and Legal Guidelines for Corporate Training
Sharon Shrock, William Coscarelli
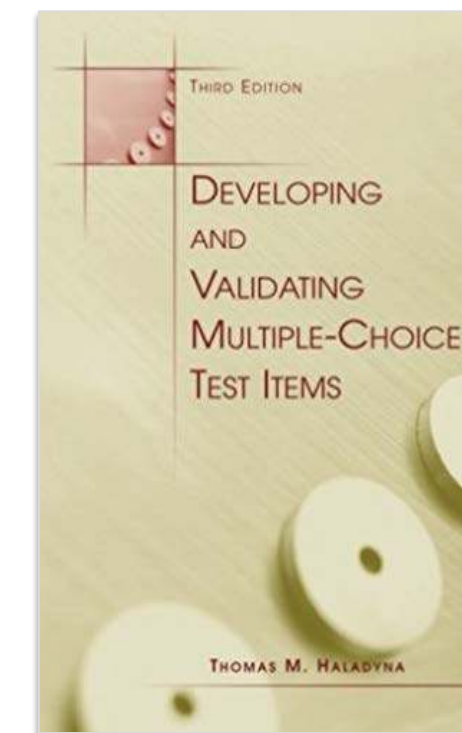
Developing and Validating Multiple-Choice Test Items
Thomas Haladyna
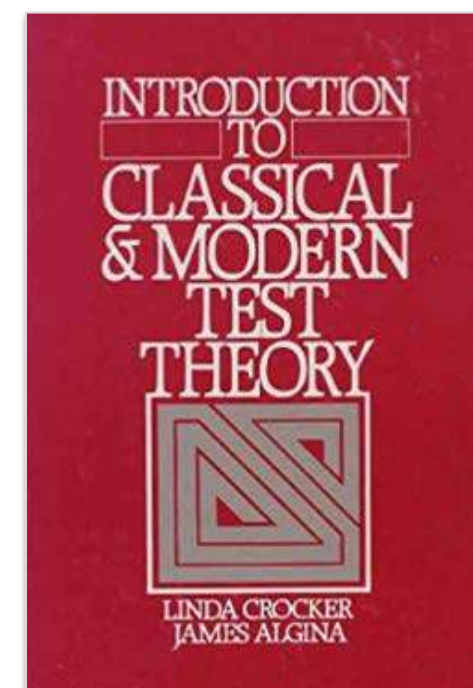
Introduction to Classical and Modern Test Theory
Linda Crocker, James Algina

Questionmark
powered by Learnosity

Useful resources

www.onetonline.org

- Database of standardized and occupation-specific descriptors occupations
- Example profile:
  - onetonline.org/link/summary/13-1031.00

Useful resources

Check out our website to find the latest:

- Reports
- Webinars
- Blog Articles
- Podcast Episodes
- And more!

Visit **questionmark.com**

Report   October 2022

The Test Fraud Fallacy

Revealing the true nature of workplace test cheating and why it could happen to you.

Questionmark
powered by Learnosity

Unlocking the Potential of Assessments

Hosted by
John Kleeman

Questionmark
powered by Learnosity

Questionmark
powered by **Learnosity**
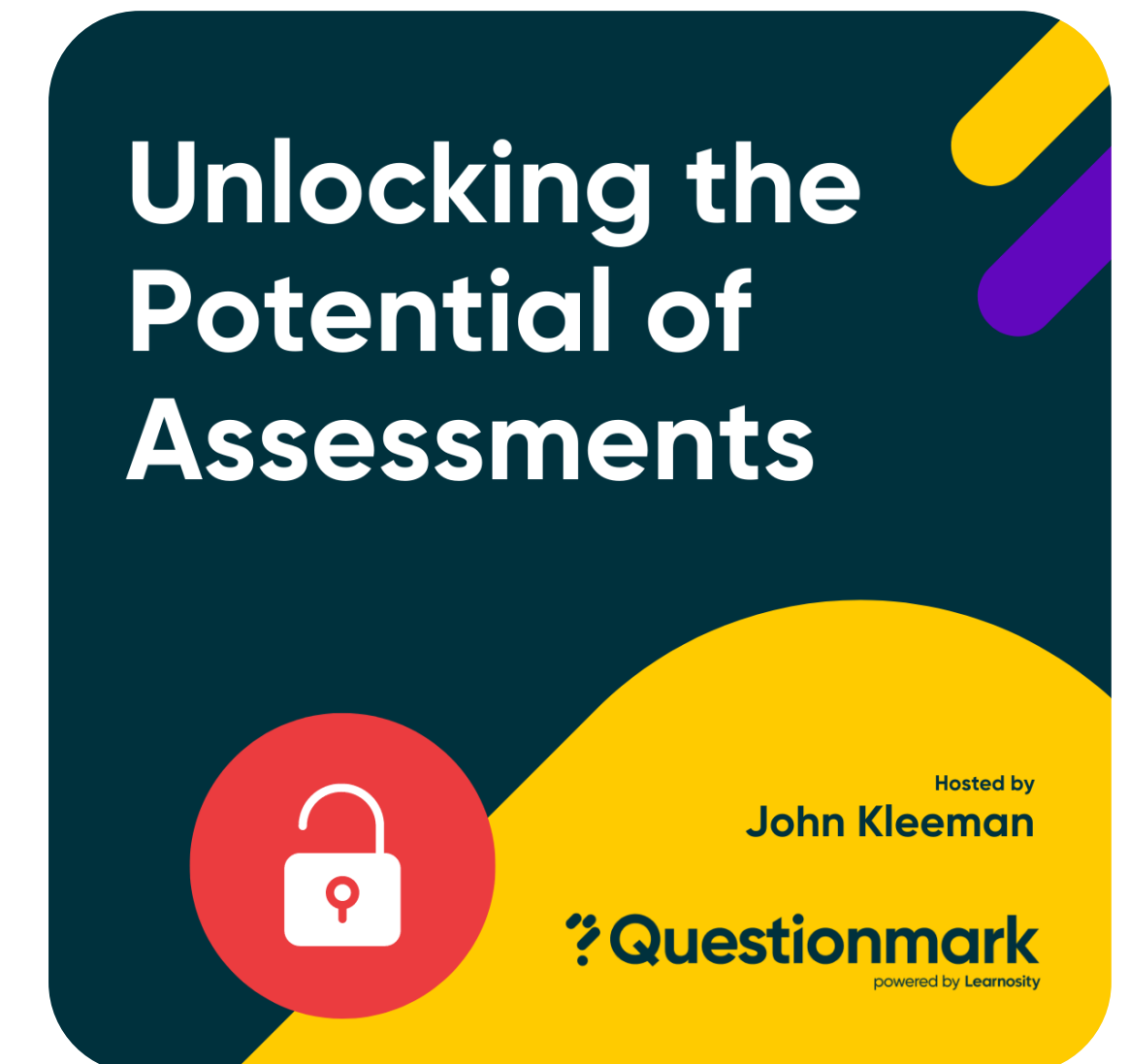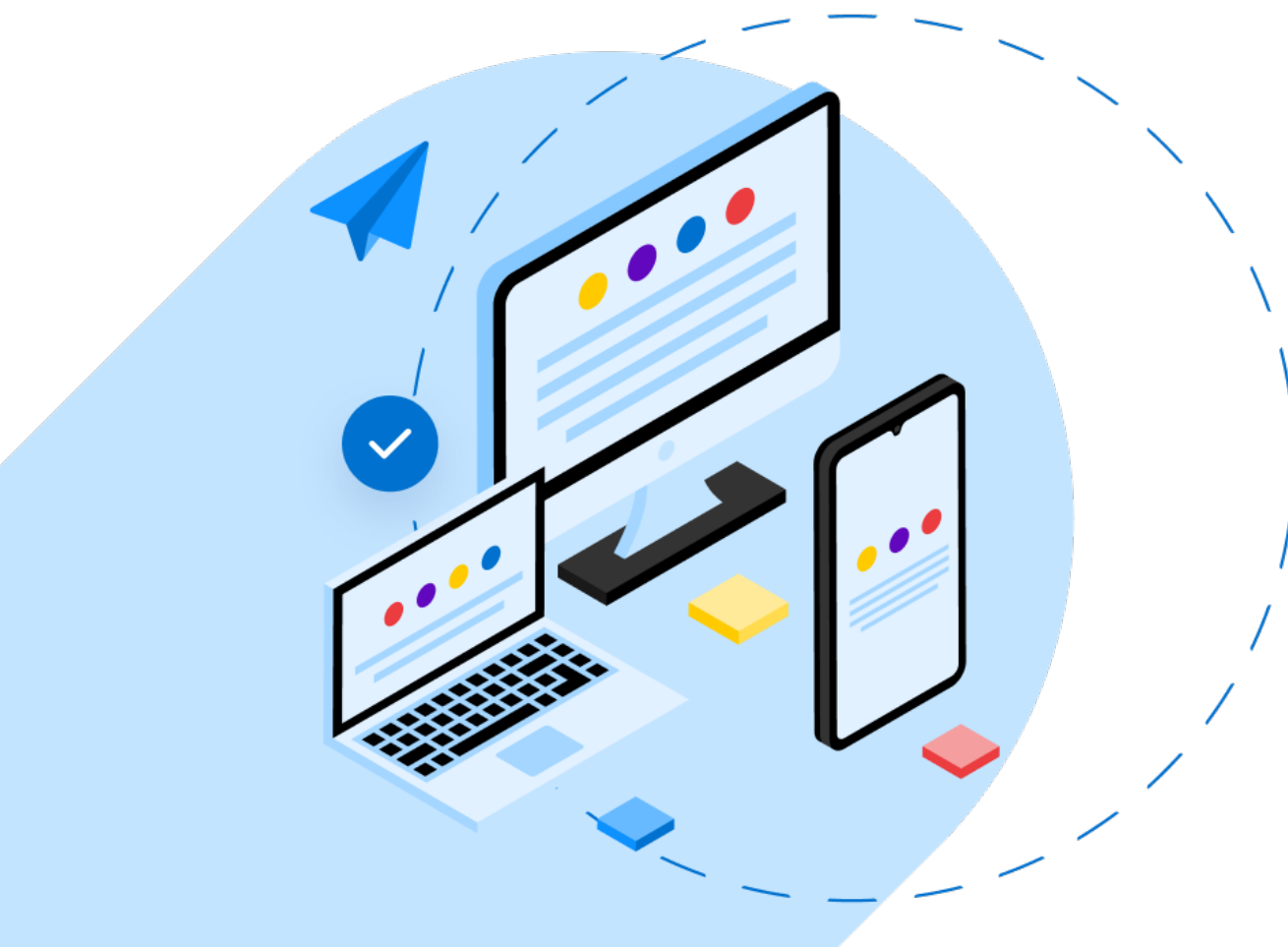
Thank you for attending!

We hope to see you at a future webinar

Questionmark

powered by **Learnosity**

questionmark.com