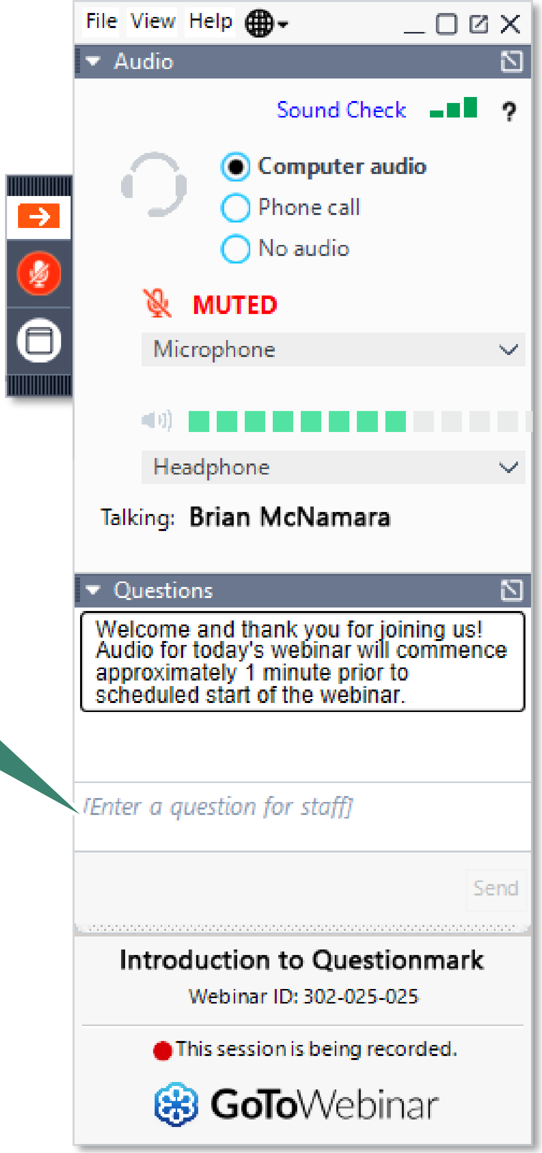
question mark

— a Learnosity company —

10 Quick Tips to Improve your Tests & Exams

Brian McNamara
Product Manager for Customer Engagement
Questionmark

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.

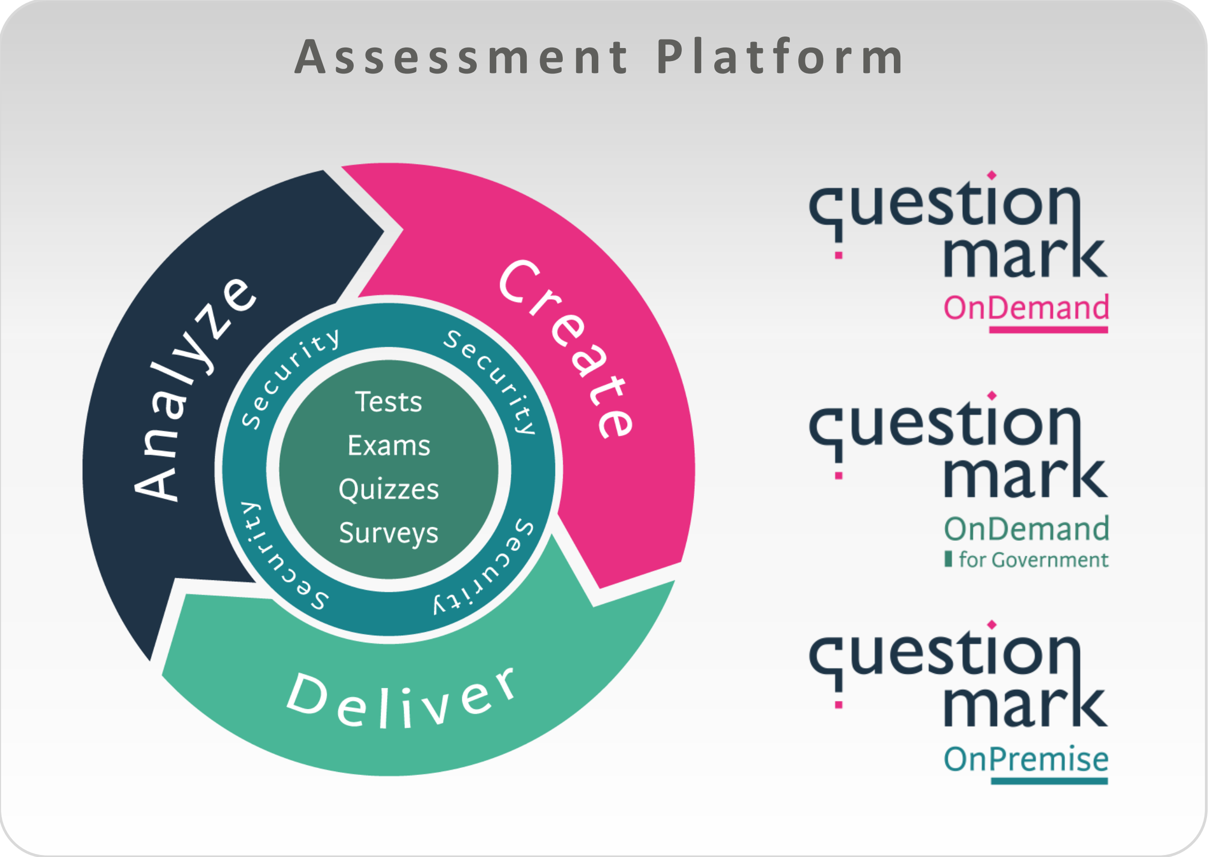
GoToWebinar

Watch for an email after the webinar to:

- Download slides (PDF)
- View a recording

About Questionmark

- Measure knowledge, skills and abilities securely
  - Assessment platform
  - Proctoring solutions
  - Assessment content
- ISO/IEC 27001 Certified
- Founded in 1988
- Part of the Learnosity Group



SAP Platinum Partner

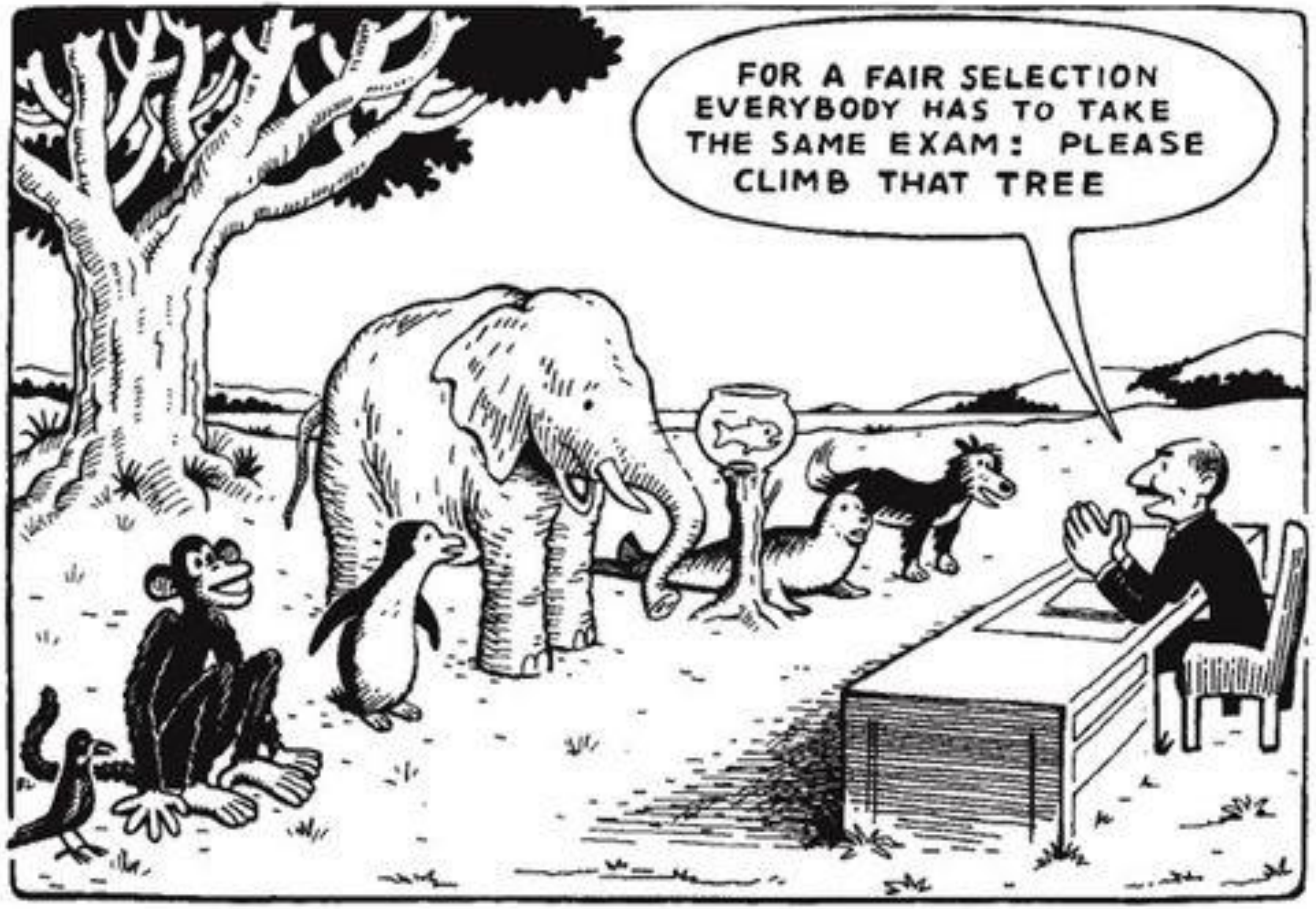
Microsoft Partner
Gold Independent Software Vendor (ISV)

bsi. ISO/IEC 27001 Information Security Management

GSA Contract Holder
Contract GS-35F-0380Y

Crown Commercial Service Supplier

What is a ‘good’ test?



FOR A FAIR SELECTION EVERYBODY HAS TO TAKE THE SAME EXAM: PLEASE CLIMB THAT TREE

Is this a good test?

question mark

— a Learnosity company —

Quick Poll ☑

Does the cartoon show a good test?

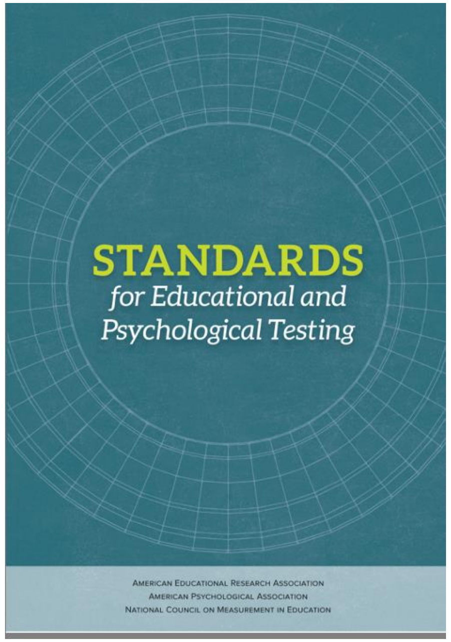
- No, it looks very unfair
- Yes, it looks a good test
- It depends what you are using the test for

This webinar: 10 questions we will answer

1. What makes a “good” test or exam?
2. How do I decide which areas to cover in a test?
3. How many questions should I include for each objective?
4. Is it safe and defensible to select questions at random?
5. Should my test/exam be open book or closed book?
6. What time limit should I set?
7. How can I work out a defensible cut score / pass mark?
8. What happens if some topics/questions are “must get right”?
9. What feedback should I give?
10. What are good resources to find out more?

Two good sources for a deeper dive

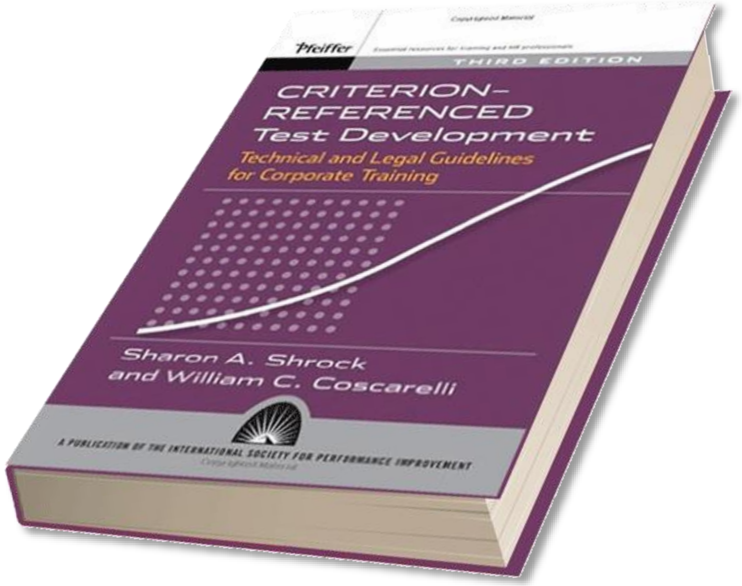
AERA/APA/NCME Standards
“The Standards”



STANDARDS
for Educational and
Psychological Testing

American Educational Research Association
American Psychological Association
National Council on Measurement in Education

Shrock and Coscarelli
Criterion Referenced Test Development



Pfeiffer

THIRD EDITION

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

Sharon A. Shrock
and William C. Coscarelli

A PUBLICATION OF THE INTERNATIONAL SOCIETY FOR PERFORMANCE IMPROVEMENT

1. What makes a “good” test or exam?

The Standards suggest three foundations for 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

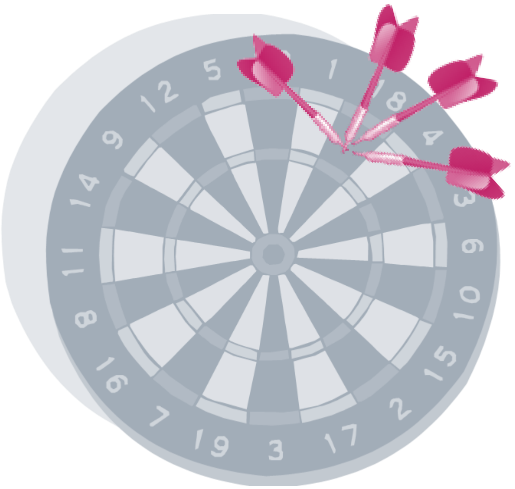
Fairness

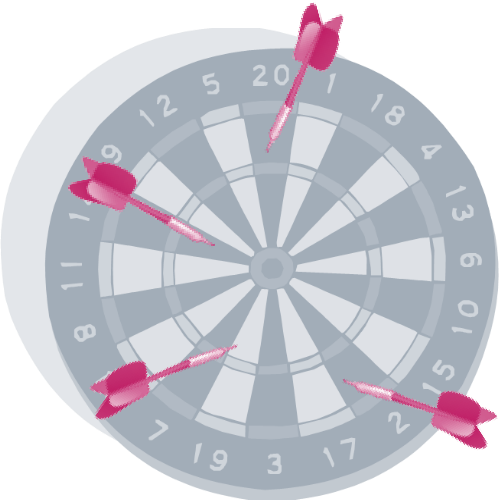
- Fair and equitable treatment of all individuals in the intended population of test-takers
- Does not advantage or disadvantage individuals because of characteristics irrelevant to the construct being measured

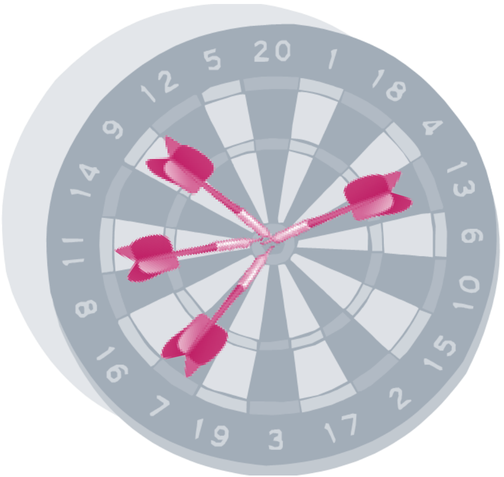
Validity and Reliability

Reliable:
- Dependable, repeatable, consistent

Valid:
- Measures appropriate knowledge and skills


Reliable but not Valid


Not Reliable, not Valid


Reliable and Valid

Three common approaches to Validity

Content validity

- Whether assessment content and composition is appropriate given what is being measured, e.g. does test cover knowledge/skills required to do a job

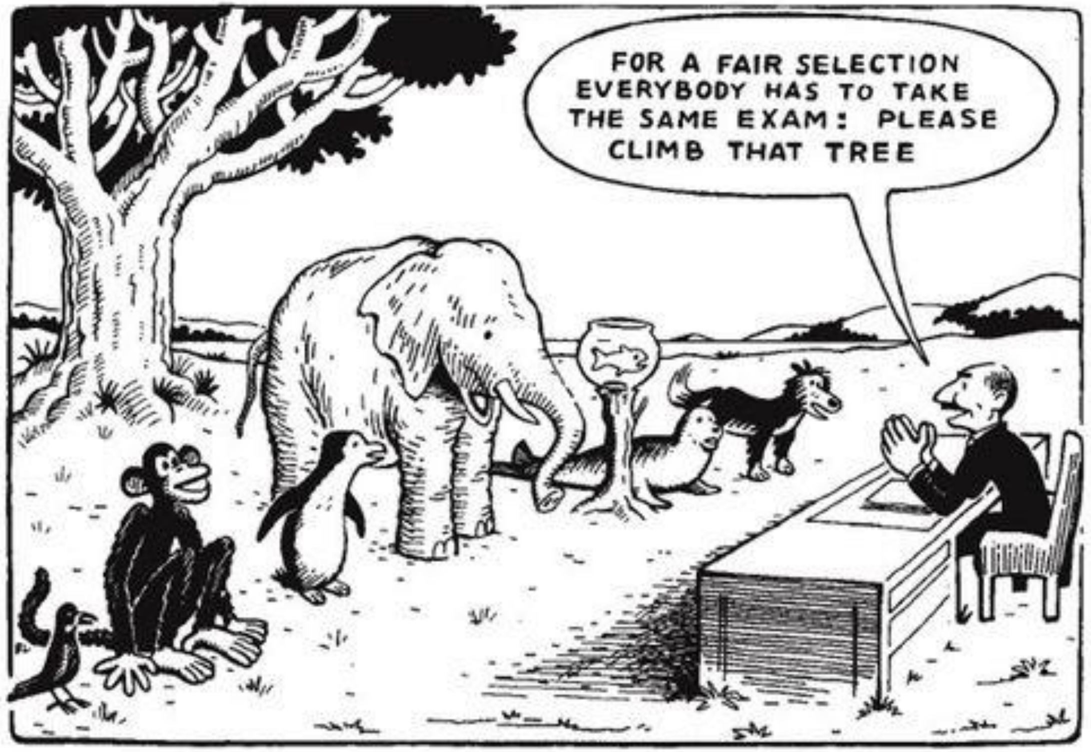
Criterion validity

- Whether test-taker assessment scores are related to other measures, e.g. do exam scores predict future performance?

Face validity

- Whether appears valid to test-takers and stakeholders

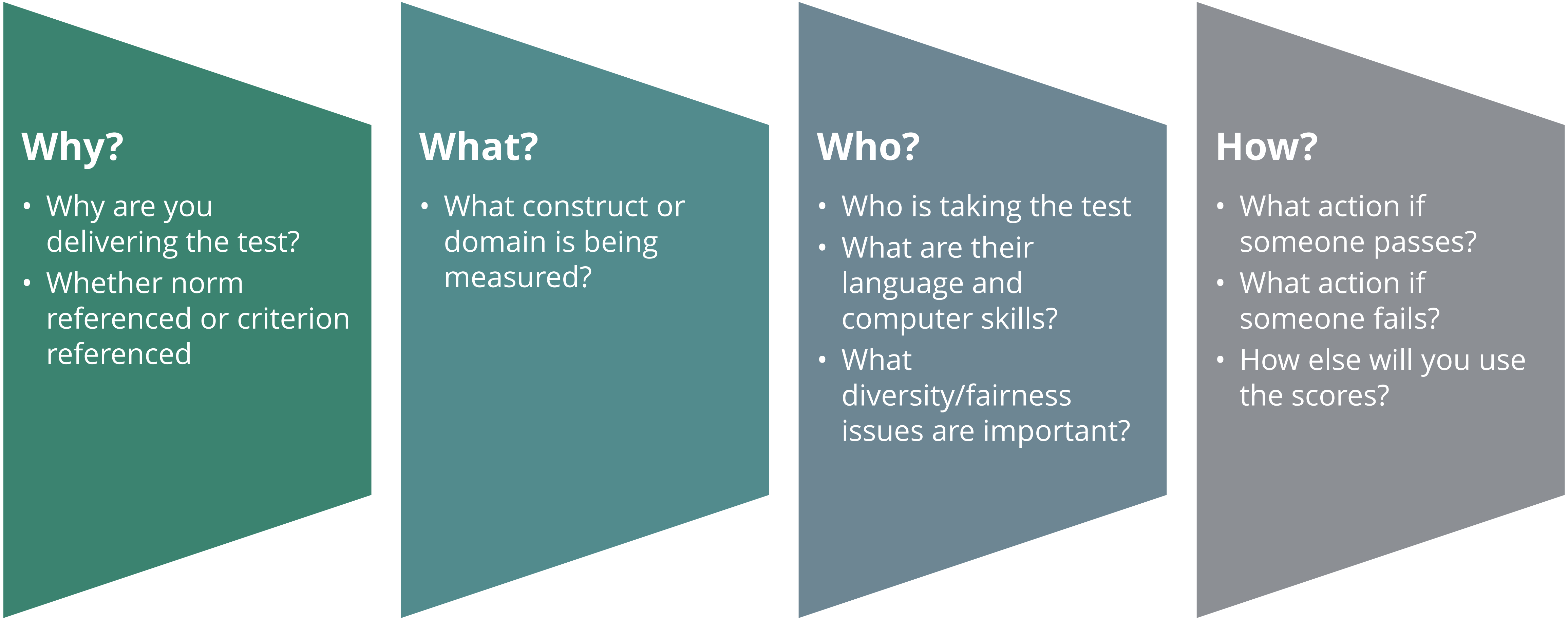
Is this a good test?


FOR A FAIR SELECTION EVERYBODY HAS TO TAKE THE SAME EXAM: PLEASE CLIMB THAT TREE

- Cartoon used to illustrate unfairness of standardized tests in education.
- However whether valid, reliable and fair depends on purpose
- For most purposes, it would be unfair. To recruit for a fruit-picking job in trees, might be useful part of selection process

2. How do I decide which areas to cover in a test?

Start with 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?

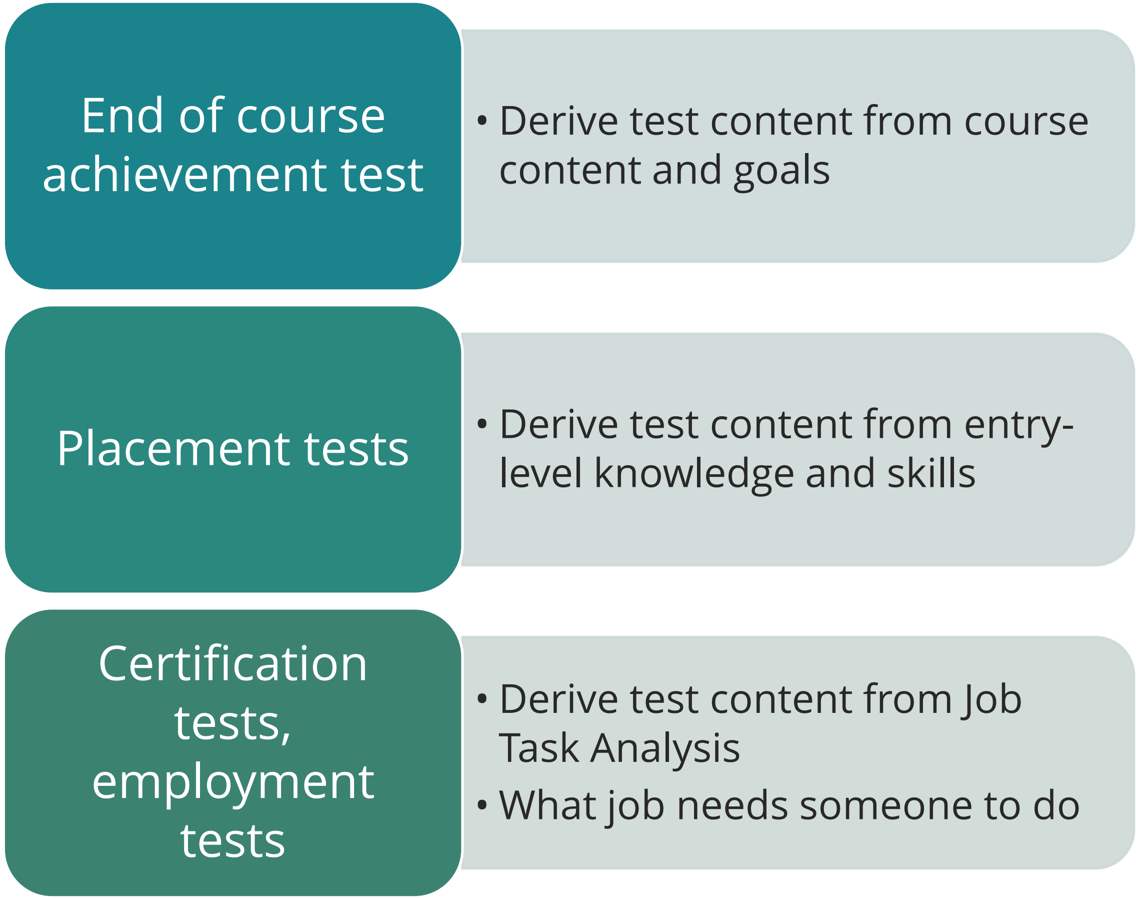
Who?

- 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?

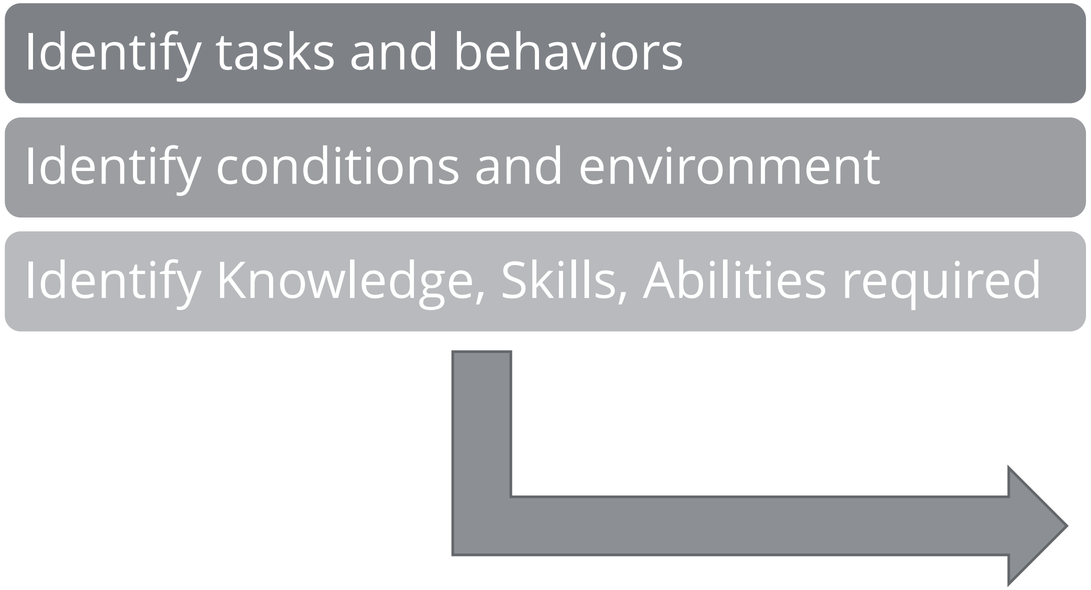
Determine content of the test based on the purpose



End of course achievement test	• Derive test content from course content and goals
Placement tests	• Derive test content from entry-level knowledge and skills
Certification tests, employment tests	• Derive test content from Job Task Analysis • What job needs someone to do

- Develop a test blueprint (AKA “test content outline”)
- Covers what is included and excluded
- Often a series of objectives and a weighting
- May include key knowledge or skills areas

Job Task Analysis



Identify tasks and behaviors

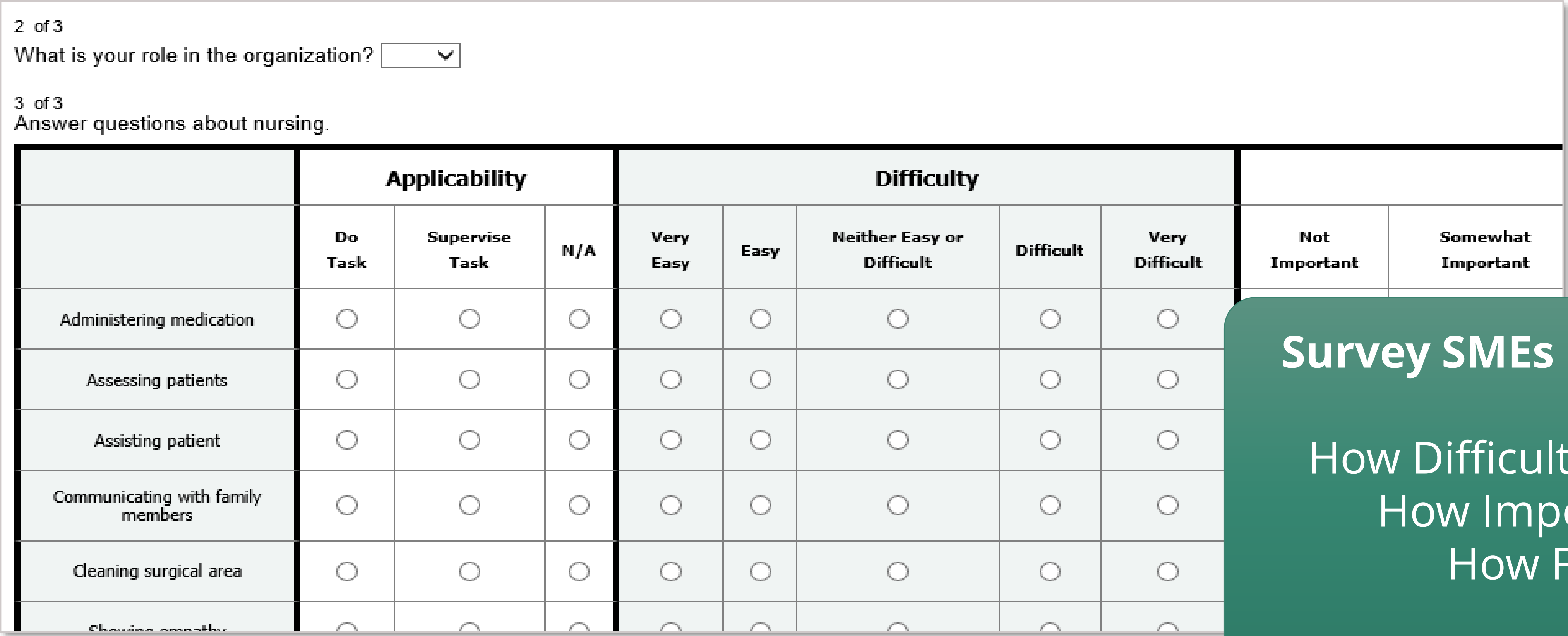
Identify conditions and environment

Identify Knowledge, Skills, Abilities required

Methods

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

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



2 of 3

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	Not Important	Somewhat Important
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?

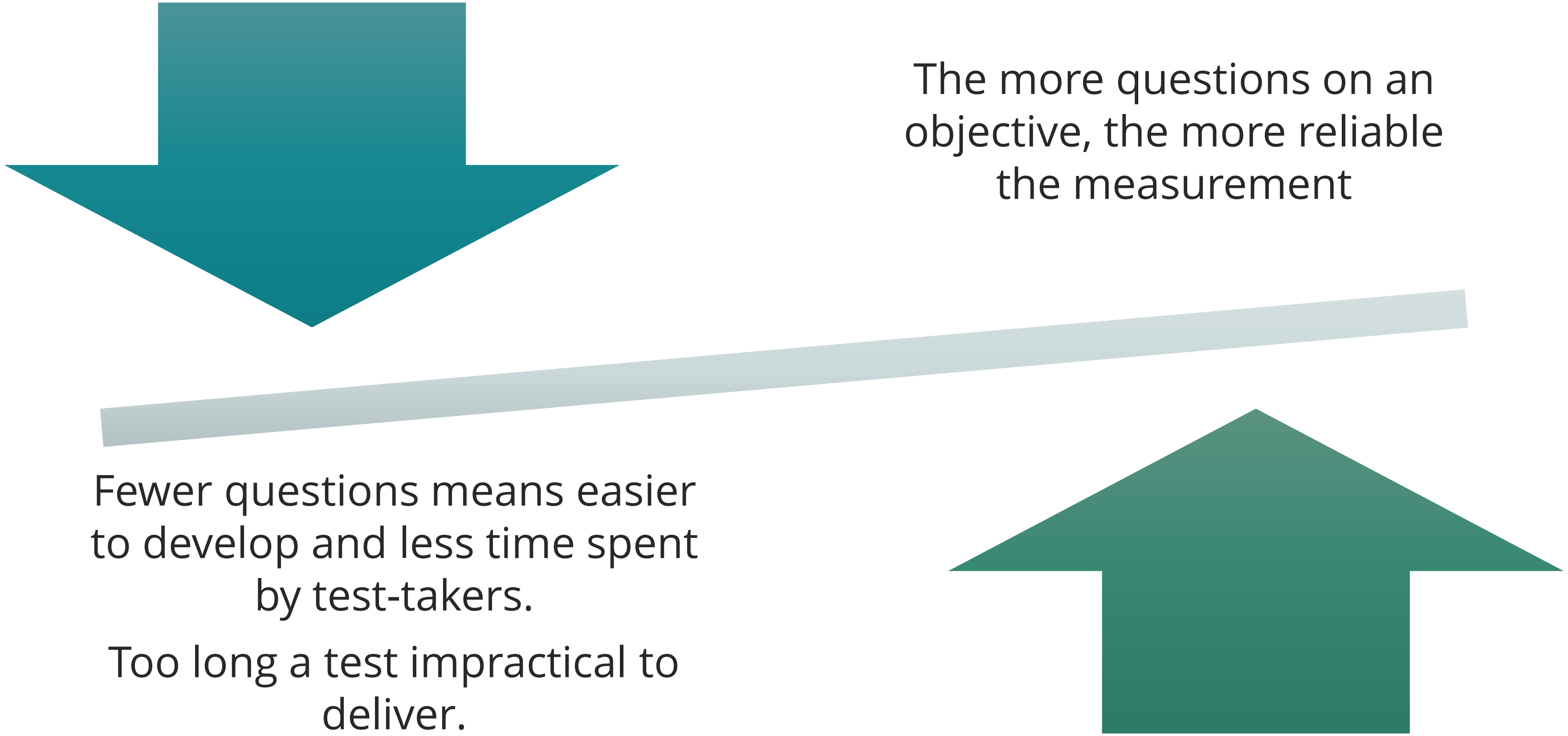
- Job Task Analysis to determine and **validate** content of test
- Test Blueprint built from JTA, which drives test item development.

questionmark

— a Learnosity company —

3. How many questions should I include for each objective?

A balance is needed



The more questions on an objective, the more reliable the measurement

Fewer questions means easier to develop and less time spent by test-takers.

Too long a test impractical to deliver.

question mark

Advice on number of questions per objective:

Research evidence suggest 4-6 items generally per objective

- More adds increasingly less value
- Fewer risks not testing objective properly

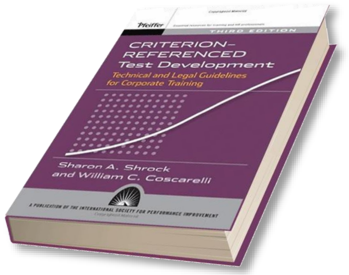
More questions needed for

- Critical objectives (e.g. health and safety)
- Large domain covered by objective (e.g. “Given access to manuals, diagnose the source of a radiation leak in a nuclear reactor”)

Less questions needed for

- Smaller domain (something very specific e.g. “List the 6 steps required to make a milkshake on a specific machine”)
- If objectives related and so doing well in one likely will mean doing well in the other

Guidance from Shrock & Coscarelli


Pfeiffer
THIRD EDITION
CRITERION-REFERENCED
Test Development
Technical and Legal Guidelines for Corporate Training
Sharon A. Shrock
and William C. Coscarelli

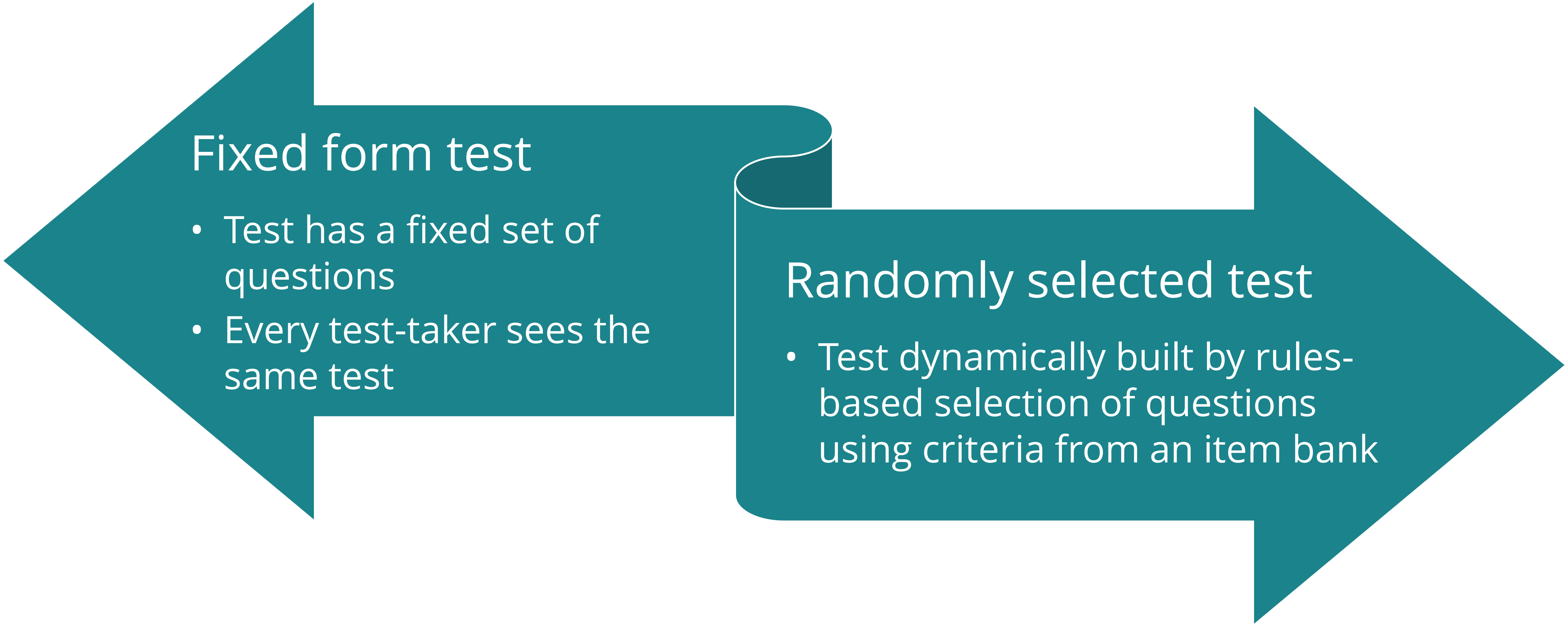
Criticality?	Domain size?	Related?	# questions
Critical	From a large domain	Unrelated	10-20
		Related	10
	From a small domain	Unrelated	5-10
		Related	5
Not critical	From a large domain	Unrelated	6
		Related	4
	From a small domain	Unrelated	2
		Related	1

question mark

— a Learnosity company —

4. Is it safe and defensible to select questions at random?

Random vs fixed form



Fixed form test

- Test has a fixed set of questions
- Every test-taker sees the same test

Randomly selected test

- Test dynamically built by rules-based selection of questions using criteria from an item bank

Which is best?

Advantages of fixed form

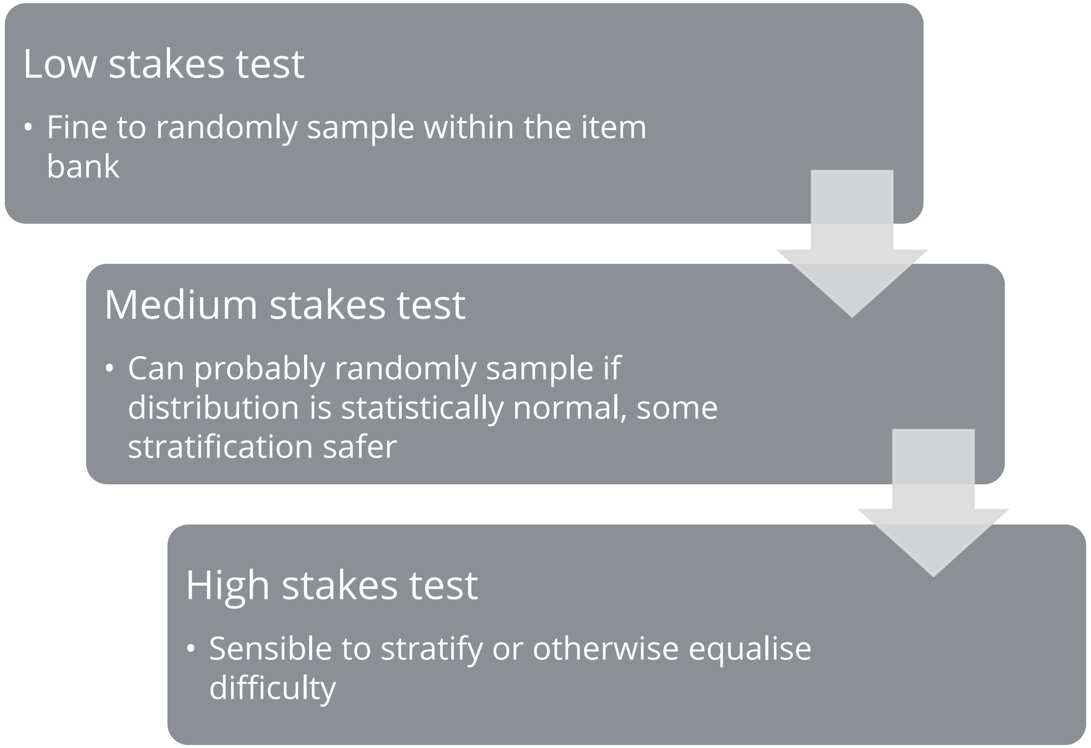
- Each test-taker gets same test which is fairer (no risk of someone getting easier or harder test)
- Need to write less questions
- Ensure no questions give the answer to others
- Gets more complicated if multiple forms needed

Advantages of random selection

- Reduces risk of cheating
- Reduces item exposure
- Easily retire or add individual questions without impacting test
- Easier to deliver test on demand (all the time not just one fixed timeslot)

How can you deal with varying difficulties?

Shrock and Coscarelli



Low stakes test

- Fine to randomly sample within the item bank

Medium stakes test

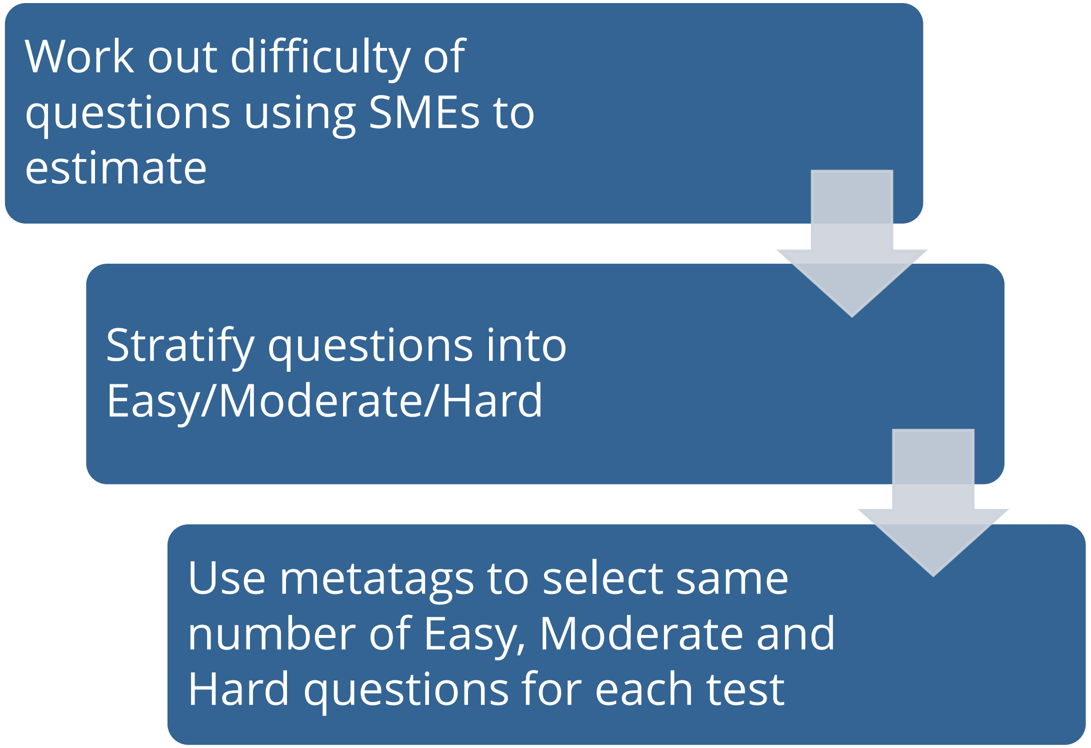
- Can probably randomly sample if distribution is statistically normal, some stratification safer

High stakes test

- Sensible to stratify or otherwise equalise difficulty

The problem of saltatory cut-score: some issues and recommendations for applying the Angoff to test Item Banks

Case study: US Coast Guard approach



Work out difficulty of questions using SMEs to estimate

Stratify questions into Easy/Moderate/Hard

Use metatags to select same number of Easy, Moderate and Hard questions for each test

Randomly designed tests – how can they be fair to all

5. Should my test/exam be open book or closed book?

What is the difference?

Open book test

- Allows test-takers to have reference books, notes or tools available whilst taking the test

Closed book test

- Requires test-takers to answer all questions from their own knowledge without access to reference resources

Arguments for open book tests

Closer to performance environment

- In real work settings, we can look things up. Why not also in exams?

Reduces test anxiety

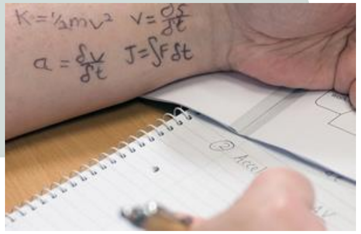
- Evidence suggests open book exams less stressful than closed book exams

Encourages testing higher level thinking skills

- More relevant for most work skills

Reduces cheating

- No longer illegal to bring in notes

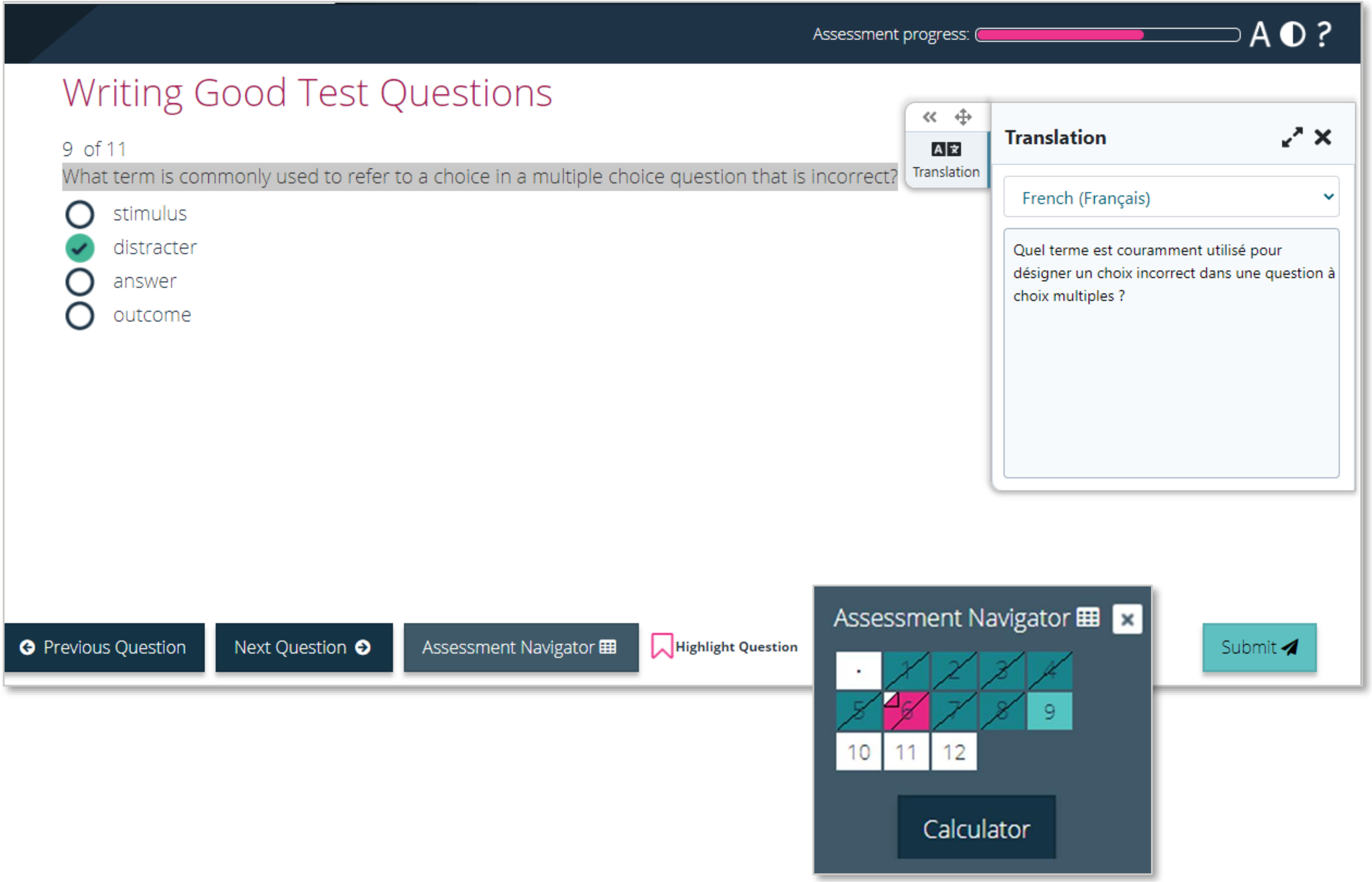


Arguments in favor of closed book tests

More conventional	• May also give more face validity because people are used to them
Important to know key knowledge in most job roles	• If a fire starts, you don’t want to have to Google something to remember what to do
Closed book tests easier to create	• Open book tests need a little more imagination in item writers
Fairness considerations…	• Could some test-takers not afford expensive text books? • Might some people bring in someone else’s notes? • Risk of answers being available on the web

In-exam tools and resources

- Provide a standard set of resources to all test-takers – for example:
  - Reference materials
  - Calculator or other tools
  - Machine Translation (new option!)



Writing Good Test Questions

9 of 11

What term is commonly used to refer to a choice in a multiple choice question that is incorrect?

○ stimulus
● distracter
○ answer
○ outcome

Translation

French (Français)

Quel terme est couramment utilisé pour désigner un choix incorrect dans une question à choix multiples ?

questionmark

— a Learnosity company —

Quick Poll ☑

Within your organization, do you use open or closed book tests?

- All closed book
- Mostly closed book
- Some of each
- Mostly open book
- All open book

6. What time limit should I set?

What is the purpose of the test/exam?

Power tests

- Measures knowledge / skill of test-taker
- Most common type of test
- Most people should have enough time to answer all questions

Speed tests

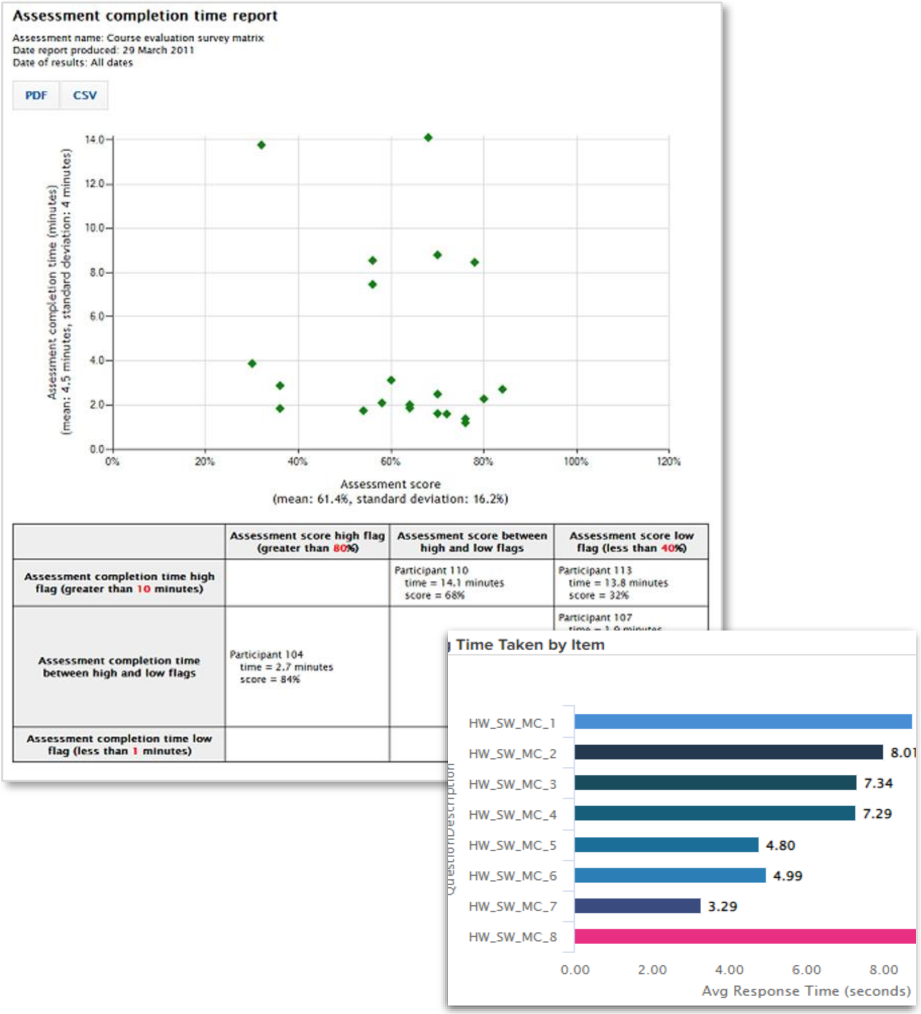
- Measure speed of test-taker in making responses
- Useful when fast speed an important part of job requirements
- In speed tests, many people may not finish all items

Allocating sufficient time for the test

- Ensure that the time limit doesn't start until
  - Instructions given
  - Any practice items taken
  - Any demographic information provided

  (In Questionmark, use untimed block)

- Seeing how long people take in pilot best way to work out required time
- Monitor actual time taken by test-takers to check remains reasonable



Assessment completion time report

Assessment name: Course evaluation survey matrix
Date report produced: 29 March 2011
Date of results: All dates

PDF
CSV
Assessment completion time (minutes)
(mean: 4.5 minutes, standard deviation: 4 minutes)
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0%
20%
40%
60%
80%
100%
120%
Assessment score
(mean: 61.4%, standard deviation: 16.2%)

	Assessment score high flag (greater than 80%)	Assessment score between high and low flags	Assessment score low flag (less than 40%)

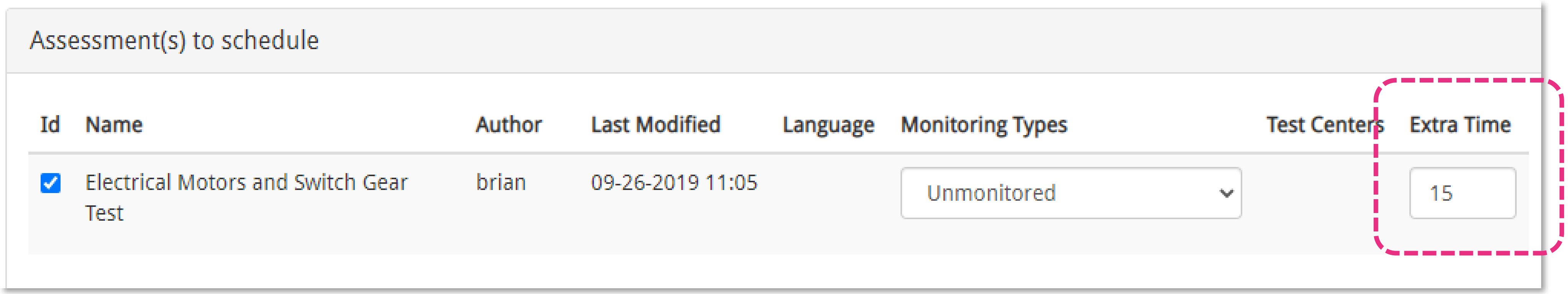
Assessment completion time high flag (greater than 10 minutes)		Participant 110
		time = 14.1 minutes
		score = 68%
			Participant 113
			time = 13.8 minutes
			score = 32%

Assessment completion time between high and low flags	Participant 104
	time = 2.7 minutes
	score = 84%
			Participant 107
			time = 1.9 minutes

Assessment completion time low flag (less than 1 minutes)			

Extra time

- Give extra time
  - Common to give extra time as accommodation for some special needs
  - Extra time also given for linguistic reasons (taking assessment in second language)
  - Ideally base the extra time on piloting (not just a fixed extra %)
- Make sure test delivery system allows you to accommodate participants as needed
  - Allocate additional time for certain individuals or groups in a “schedule” or an “exception schedule”



Assessment(s) to schedule

Id	Name	Author	Last Modified	Language	Monitoring Types	Test Centers	Extra Time
☑	Electrical Motors and Switch Gear Test	brian	09-26-2019 11:05		Unmonitored		15

7. How can I work out a defensible cut score / pass mark?

(for a criterion referenced test or exam)

question mark
— a Learnosity company —

Quick Poll ☑

What is a good cut score / pass mark for a criterion referenced test?

- 70%
- 80%
- 90%
- It depends how hard the questions are

Start with consequences of mis-classification

	Fail	Pass
Competent	Error of rejection	Correct decision: test taker should pass
Not competent	Correct decision: test taker should not pass	Error of acceptance

If consequences of error of acceptance high (e.g. surgeon, pilot), set cut score high to minimize

If error of acceptance less of a concern or consequences of error of rejection high (e.g. test taker lawsuit), may consider lower cut score.

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

More info : Webinar by Questionmark customer on the Angoff method

Why use this method?

One of our customers summed it up this way:

The Angoff Method is:

- *Defensible*
- *Easy to use and implement*
- *Widely accepted*

Angoff Method Example

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

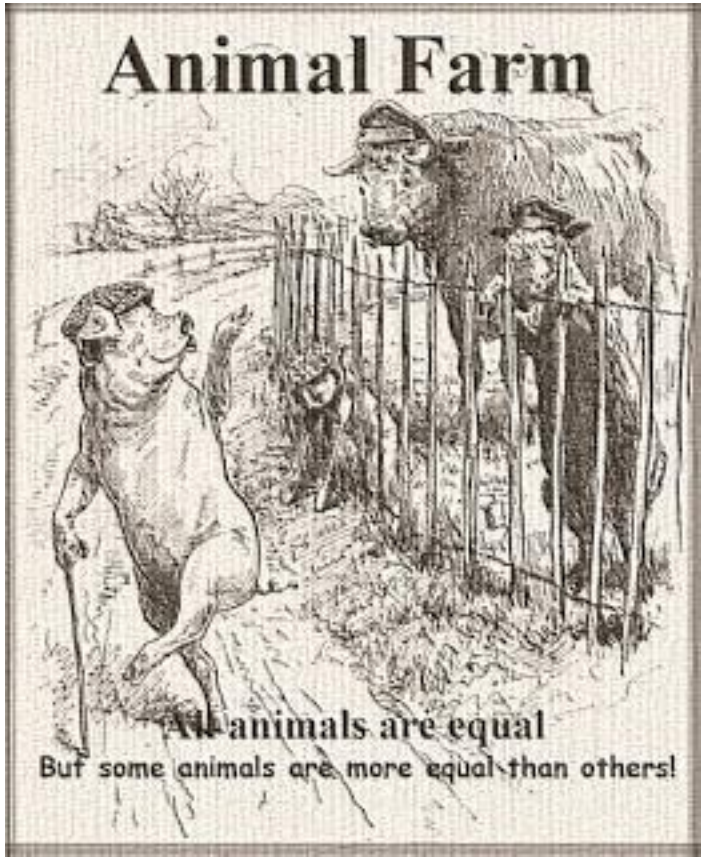
	SME A	SME B	SME C	Total
Q1	75%	75%	75%	75.00%
Q2	70%	80%	80%	76.67%
Q3	65%	75%	70%	70.00%
Q4	60%	85%	90%	78.33%
Q5	80%	80%	85%	81.67%
Q6	80%	80%	80%	80.00%
Q7	75%	80%	75%	76.67%
Q8	65%	90%	65%	73.33%
Q.....	75%	80%	75%	76.67%
Q50	65%	85%	65%	71.67%
Totals	**71%**	**81%**	**76%**	**76%**

questionmark

— a Learnosity company —

8. What happens if some topics/questions are “must get right”?

Are all your items and topics "equal"?



- Is a poor score in one item or topic made up by a good score on other items/topics?
- Or is it important that test-takers get some questions right or score well in some topics?

Golden questions or topics can be important

Are all items substitutable?

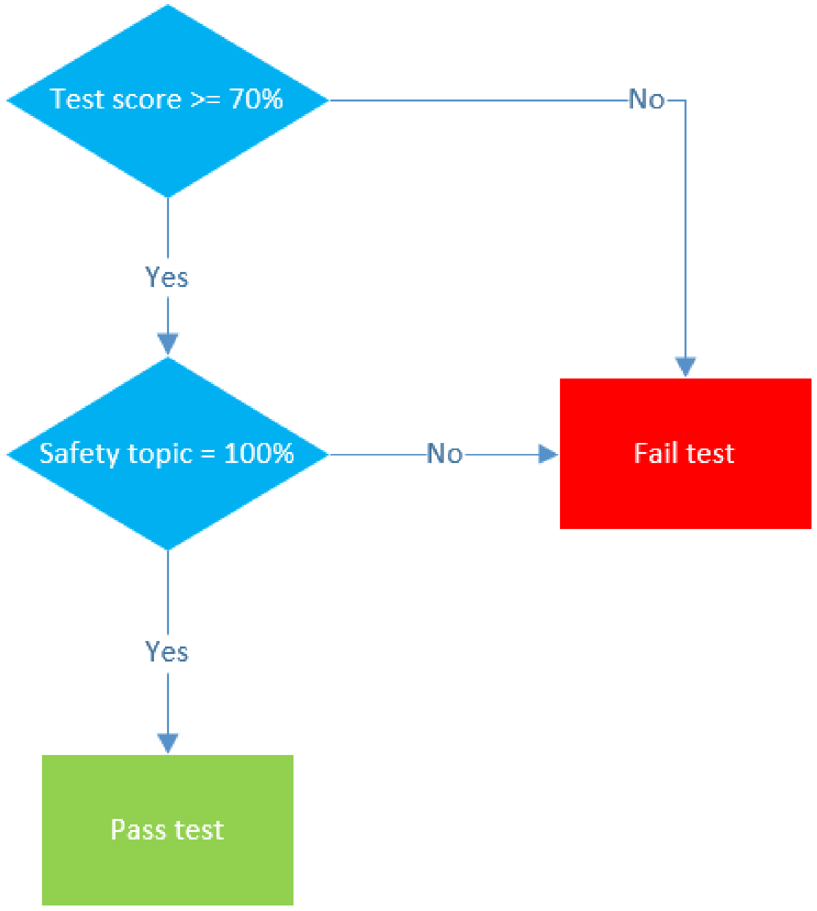
- Is a poor score in one item or topic made up by a good score on other items/topics?
- If so, you can set a single “cut score”

Some items needed for mastery?

- Sometimes, critical items or topics must be passed to show competence
  - Sometimes called “golden questions” or “golden topics”
  - Failing a safety question might mean failing the test even if all other questions are right

How do you deal with this?

- You want test-taker to pass
  - If reach the cut score
  - And meet the safety / other criteria
- In Questionmark, we have concept of a prerequisite topic, where you must pass that topic as well as meet the cut score for the test

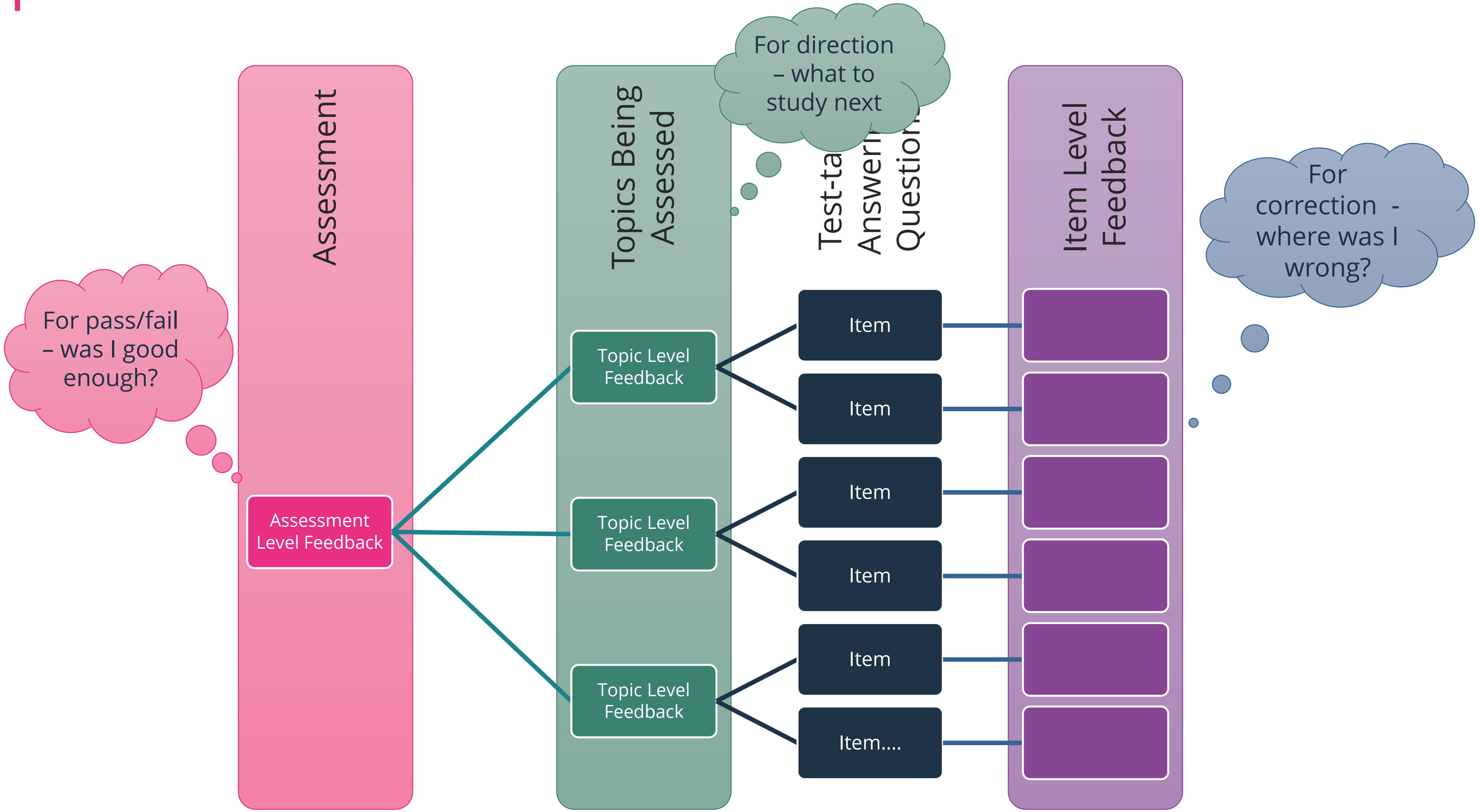


question mark

— a Learnosity company —

9. What feedback should I give?

Types of Assessment Feedback



For pass/fail – was I good enough?

Assessment

Assessment Level Feedback

Topics Being Assessed

Topic Level Feedback

Topic Level Feedback

Topic Level Feedback

For direction – what to study next

Test-ta
Answeri
Question

Item

Item

Item

Item

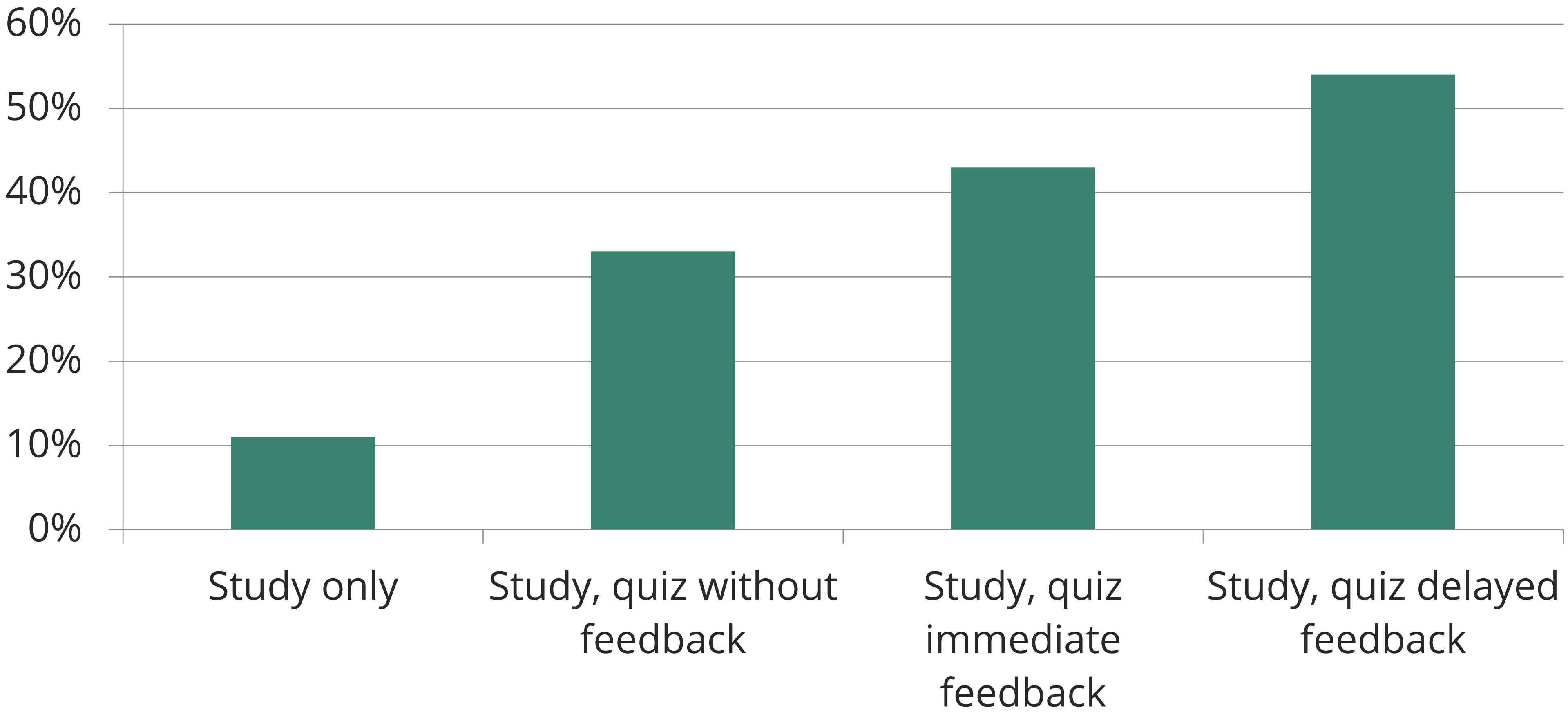
Item

Item….

Item Level Feedback

For correction - where was I wrong?

In some cases, feedback can improve understanding & retention of learning



Data showing retention after one week from Roediger & Butler : The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences 2010.

Feedback advice

Helpful in all learning contexts

- May not be appropriate in certification and some other contexts
- Most valuable to correct misconceptions

Feedback at the topic level

- Can be very helpful to direct for further study
- Most useful if topic has enough questions to be reliable (risk of small numbers of questions in a topic meaning failing or passing a topic less meaningful)

Feedback at the question level

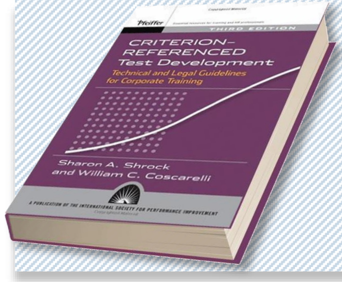
- Usually best only to give if question was wrong
- Give the correct answer
- Keep feedback short, clear and simple
- Too long feedback risks attention loss

More advice:

- Will Thalheimer, Providing Learners with Feedback
- ETS Research report on "Focus on formative feedback"

10. What are good resources to find out more?

As we already mentioned


CRITERION-REFERENCED Test Development
Technical and Legal Guidelines for Corporate Training
Sharon A. Shrock and William C. Coscarelli

Criterion-Referenced Test Development:
Technical and Legal Guidelines for Corporate Training

Sharon Shrock and William Coscarelli

Standards for Educational and Psychological Testing

AERA, APA and NCME

Some useful standards

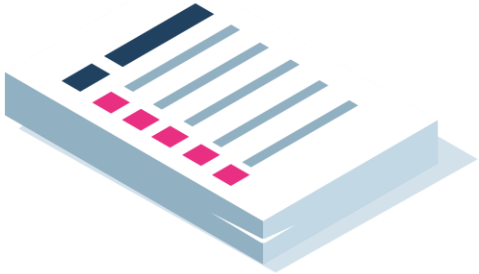
- ISO standard
  - ISO 10667: Assessment service delivery -- Procedures and methods to assess people in work and organizational settings
- Institute for Credentialing Excellence (ICE)
  - Assessment-based certificate standard
  - NCCA standards for certification programs
- International Test Commission standards
  - The ITC Guidelines on Adapting (translating) tests
  - The ITC Guidelines on the Security of Tests, Examinations and Other Assessments
  - (and others)

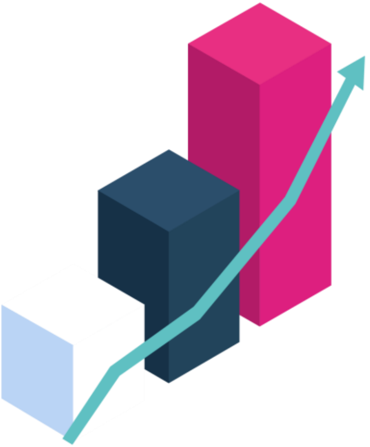
question mark

— a Learnosity company —

Your questions

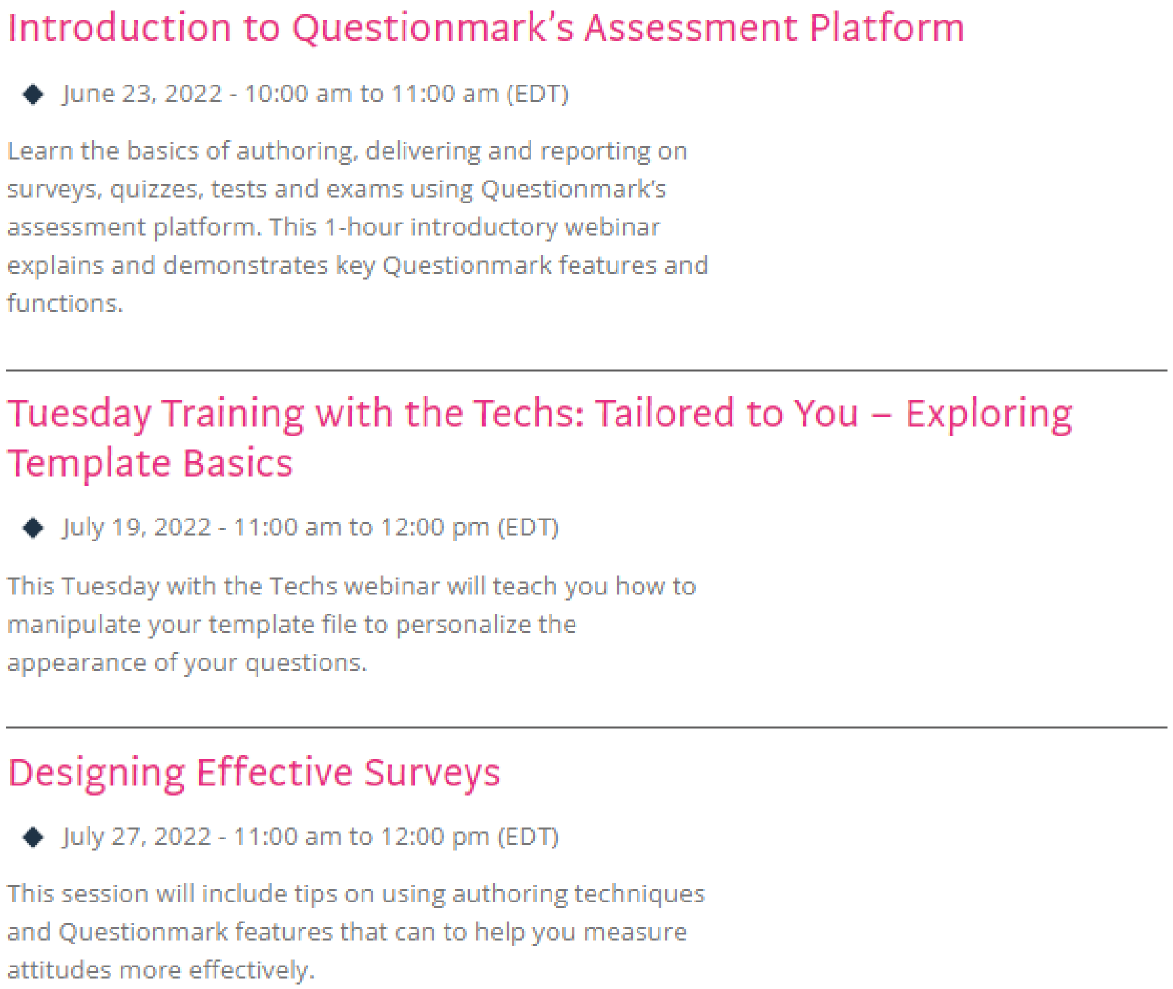
White papers, infographics, reports, eBooks and more!





www.questionmark.com/resources

Upcoming Webinars



Introduction to Questionmark's Assessment Platform

- June 23, 2022 - 10:00 am to 11:00 am (EDT)

Learn the basics of authoring, delivering and reporting on surveys, quizzes, tests and exams using Questionmark's assessment platform. This 1-hour introductory webinar explains and demonstrates key Questionmark features and functions.

Tuesday Training with the Techs: Tailored to You – Exploring Template Basics

- July 19, 2022 - 11:00 am to 12:00 pm (EDT)

This Tuesday with the Techs webinar will teach you how to manipulate your template file to personalize the appearance of your questions.

Designing Effective Surveys

- July 27, 2022 - 11:00 am to 12:00 pm (EDT)

This session will include tips on using authoring techniques and Questionmark features that can to help you measure attitudes more effectively.

questionmark

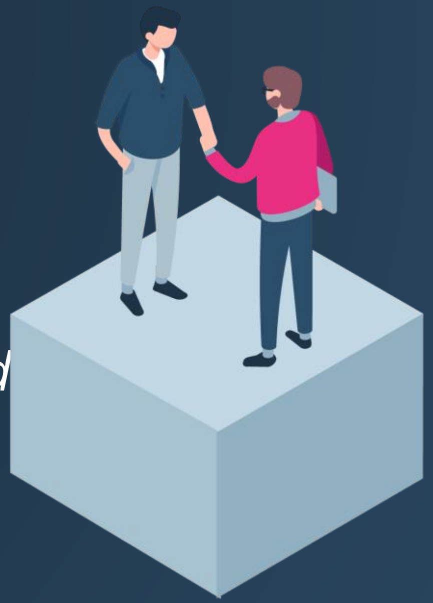
— a Learnosity company —

How to Evaluate

Request a one-on-one demo

The Questionmark team will contact you to arrange a demonstration tailored and questions

www.questionmark.com/request-demo



question mark

— a Learnosity company —

Thank you for attending!

We hope to see you at a future webinar.

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