



# Making Scores Meaningful: The Role of Standards

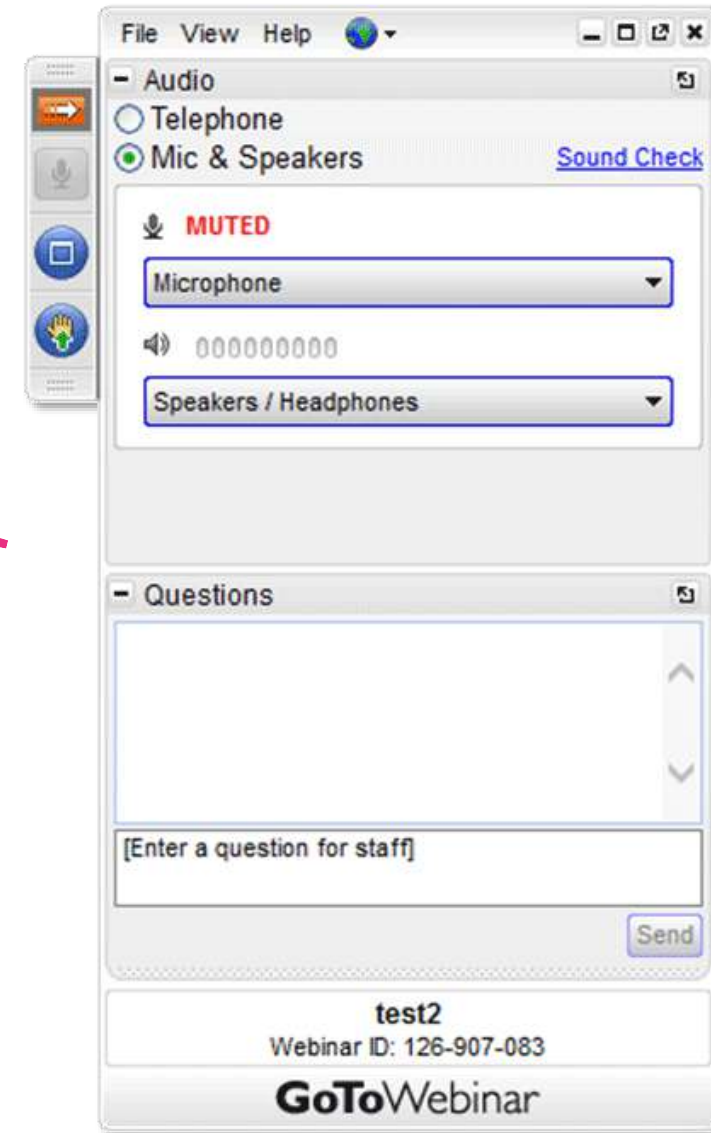
Tom Gallacher

*Psychometrics Product Manager*

To ask questions,  
use the “Questions”  
feature

**Watch for an email after the webinar:**

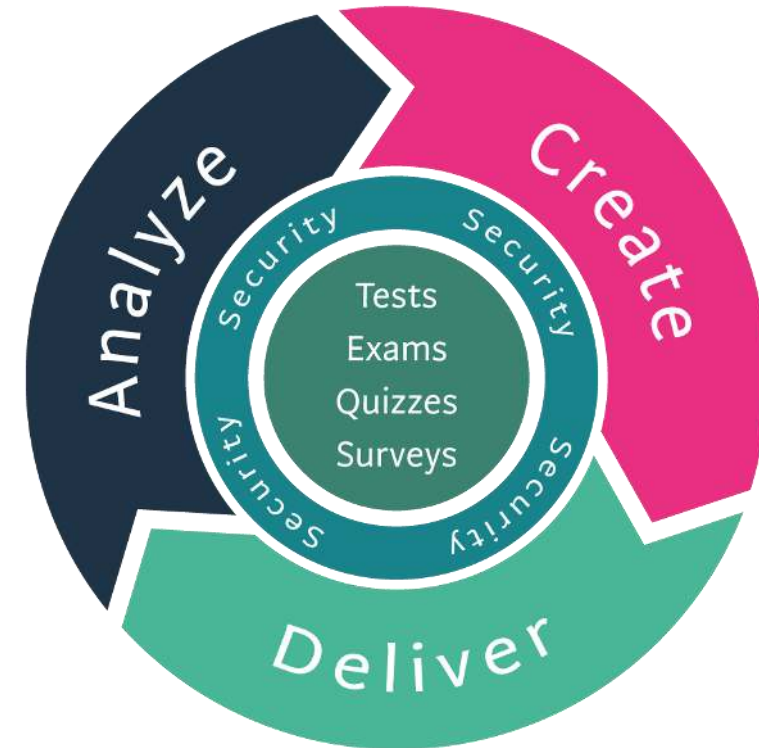
- Download slides (PDF)
- View a recording



# About Questionmark

## Background

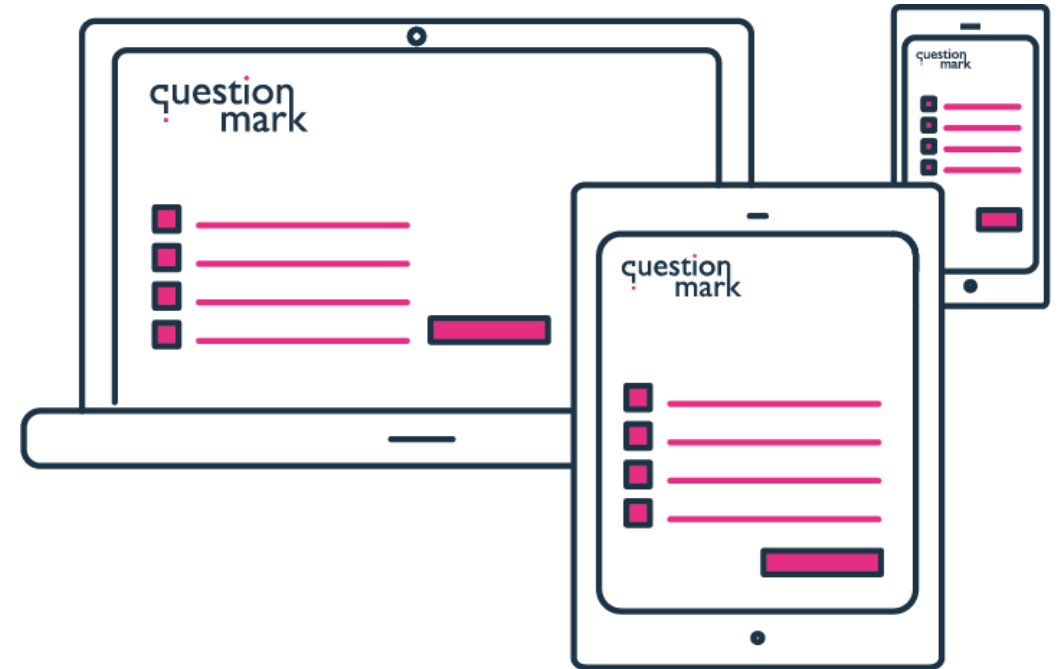
- Founded in 1988
- Assessment solutions to measure knowledge, skills, abilities and attitudes securely for certification, regulatory compliance, workforce learning, sales-force readiness and higher education
- ISO/IEC 27001 Certified (Learn more: [www.questionmark.com/trust](http://www.questionmark.com/trust))



- *Questionmark OnDemand*
- *Questionmark OnDemand for Government*
- *Questionmark OnPremise*

# Agenda

- Introduce Norm Reference and Criterion Reference through examples
- Describe what each hopes to achieve
- Contrast both to learn about the differences
- Discuss standard setting, and standard maintenance for both
- Conclude and questions



# About the presenter

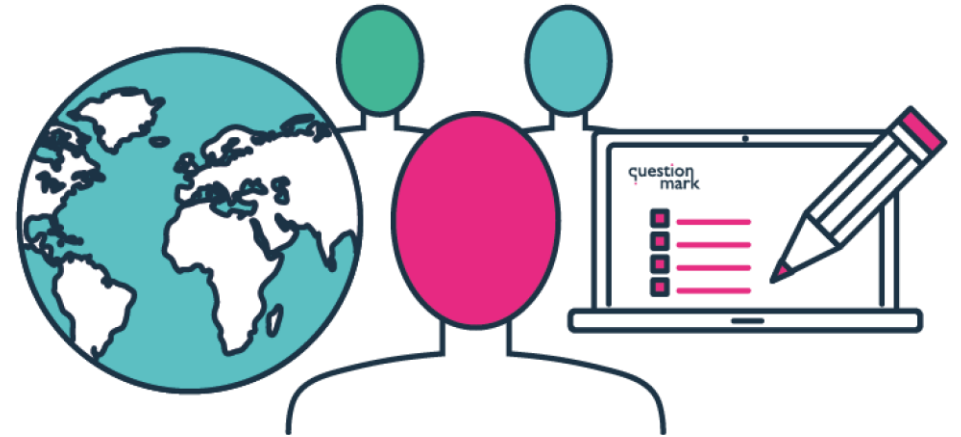
Tom Gallacher MA MSc MBPsS is the in-house psychometrician for Questionmark.

- Education in Psychology, specializing in individual differences and stats.
- Practitioner in AEA-Europe, graduate member of British Psychology society, and “fellow” of Royal Statistical Society.
- Has worked with exam boards, item writers, test publishers, and now the aspects pertaining to the platform.
- Has worked with language tests, cognitive abilities tests, situational judgement, personality, attitudinal surveys.
- Advocate of using Item Response Theory, particularly the Rasch model.
- Recycled the webinar title from his favorite book chapter!  
(<https://www.cambridgeenglish.org/Images/546163-05-scoring-validity-elliott-and-gallacher-.pdf>)



# Introduction

- Psychometrics add meaning to results:
  - Item and test quality
  - What test takers can do and what they cannot do
- Standards help the latter, they add the 'so what'
- Act like a ruler to quantify similarities and differences
- Not *necessary* in some situations, always *advantageous*



# How do they work in practice?

- Add up points on the assessment to get a raw score.
- Convert a raw score into a grade or scaled score.
- The scaled score might look completely different
- The raw score only reflects the responses to the questions, the scaled score reflects the standard.

From	To	Assessment Outcome
0%	69%	Fail
70%	100%	Pass

- Two types:
  - Criterion reference: the assessment outcome reflects a set of checklists
  - Norm reference: the assessment outcome reflects a set of other respondents

## Quick Poll

### **Which type of standard is better?**

- Criterion referenced standard
- Norm referenced standard
- Don't know about either
- Neither is better/it depends
- Need a fusion of both



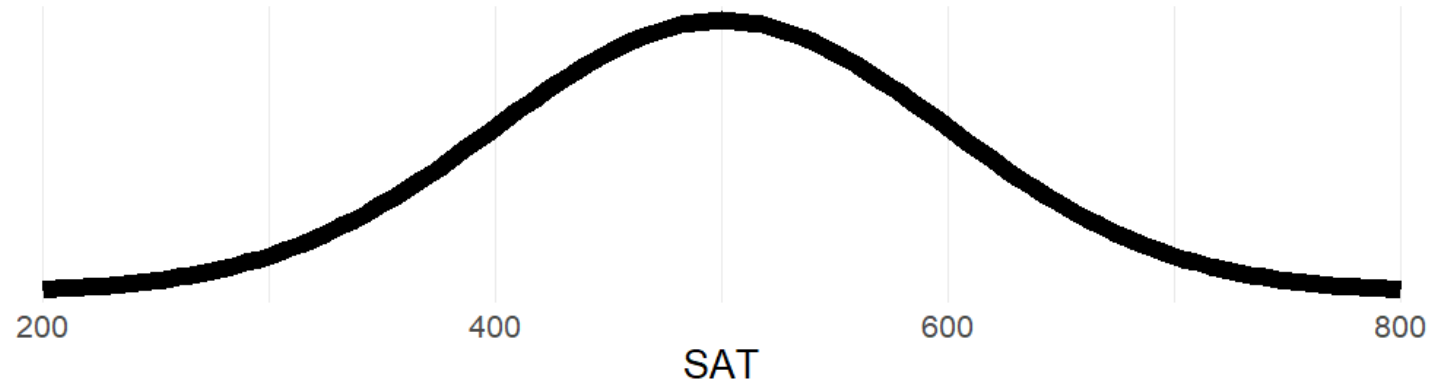
# Criterion referenced standard: driving test

- Driver and Vehicle *Standards* Agency
- Huge risk makes assessing important
- Complex and contextual skills
- Not just one dimension of performance?
- Communicating standards allows 'what gets measured gets done' positive washback
- Standards allow comparison of results across jurisdictions



# Norm referenced standard: SAT

Normal distribution of SAT Math



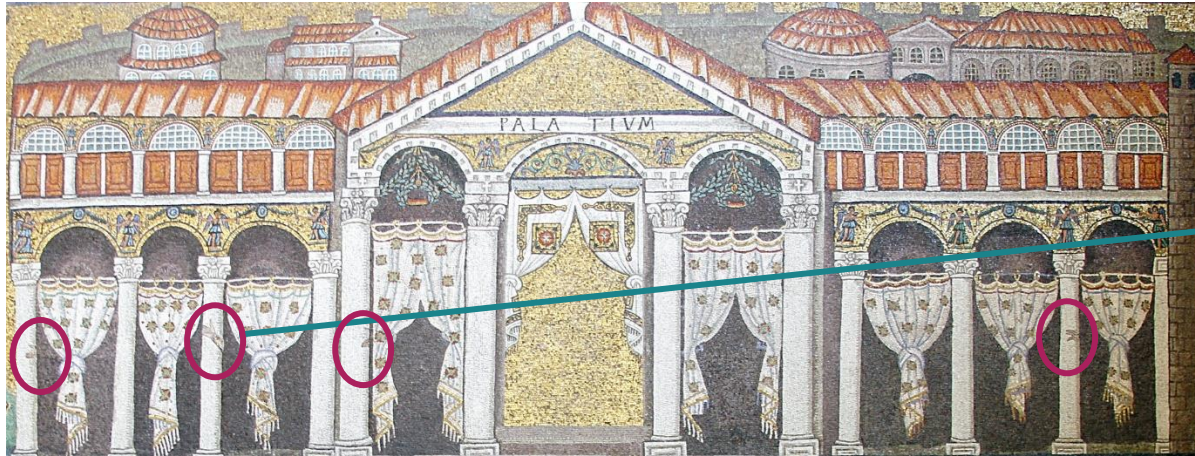
- Facilitates fine gradations rather than broad buckets
- Facilitates discussion of groups of performances
- Setting up equivalence to populational statistics is easy

# Differences and middle ground



- You might use a criterion reference: know what needs to be learnt and to what level
- You might use a norm reference: fixed percentage of test takers to pass
- There can be a middle ground, but this comes with tensions
- E.g. England's educational qualifications during COVID

# No standard is set in stone!



- The standard being maintained in stone at Theodoric's palace in Ravenna, Italy. By © José Luiz Bernardes Ribeiro, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=52650118>
- Driving standard has to keep up with new types of vehicle!
- Norm groups need refreshing to keep up to date!



# First hurdle:

setting a standard

# Setting a standard

## Criterion reference

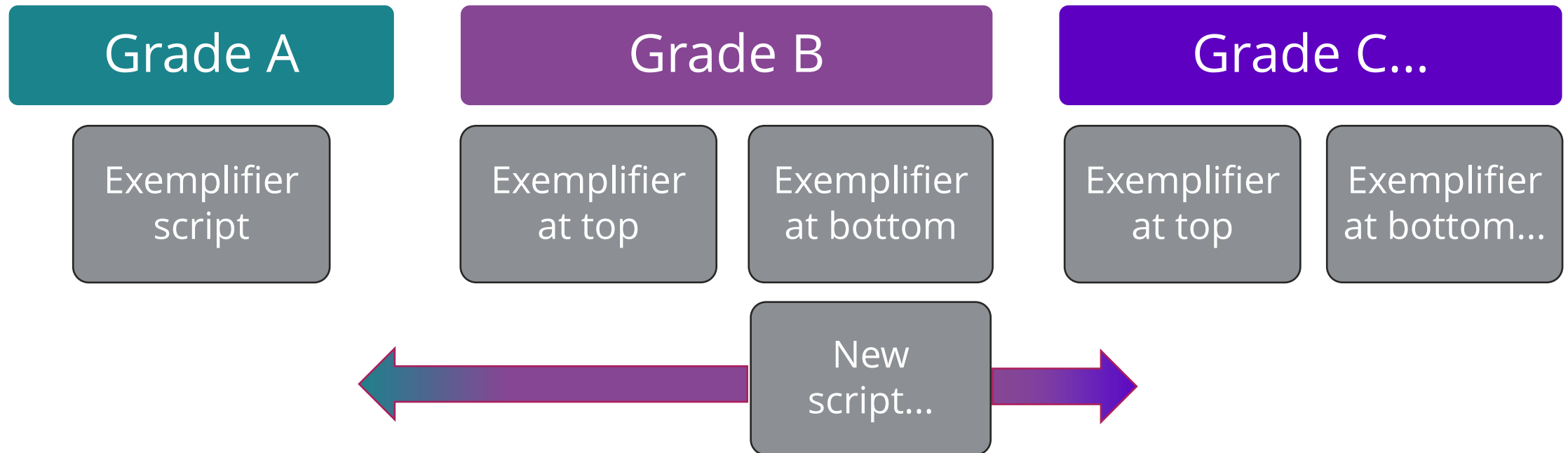
- For criterion reference: See our webinar on “Using the Angoff method to set cut scores”
  - Angoff method defines the minimal level of performance on each new test
  - Can be intensive
  - Can be done in advance of exposing the material
  - Might struggle to be flexible

## Norm reference

- For norm reference: collect performances from a large enough group
  - The vanguard cannot be given standardised results until the norms are established
  - Can be difficult to build enough momentum

# Setting a standard (third approach?)

- Requires you have some expertise of criteria, **and** some access to real performance...
- Particularly useful for essays or portfolios
- Choose a performance to exemplify grade boundaries
- Compare each new performance to these exemplifiers





# Second hurdle onwards:

maintaining a standard



# Maintaining a standard

- Monitoring results as they come in to evaluate what needs tweaking

## Criterion reference

- Angoff per set of questions?
- Equating and linking assessments?
- See our webinar “Psychometrics 101”.

## Norm reference

- Do the scores continue to meet expectations?
- Flynn effect: test taking got better over C.20<sup>th</sup>



# Item response theory!

- My personal favourite
- Knitting in new assessment content into the scale of difficulty and ability of the standard
- Can put ability estimates on a norm or criterion scale.

## Are standards worthwhile?

- Are we all *moderately* bamboozled?
- Low stakes assessment does not need a standard
- Job interview is unstandardised also

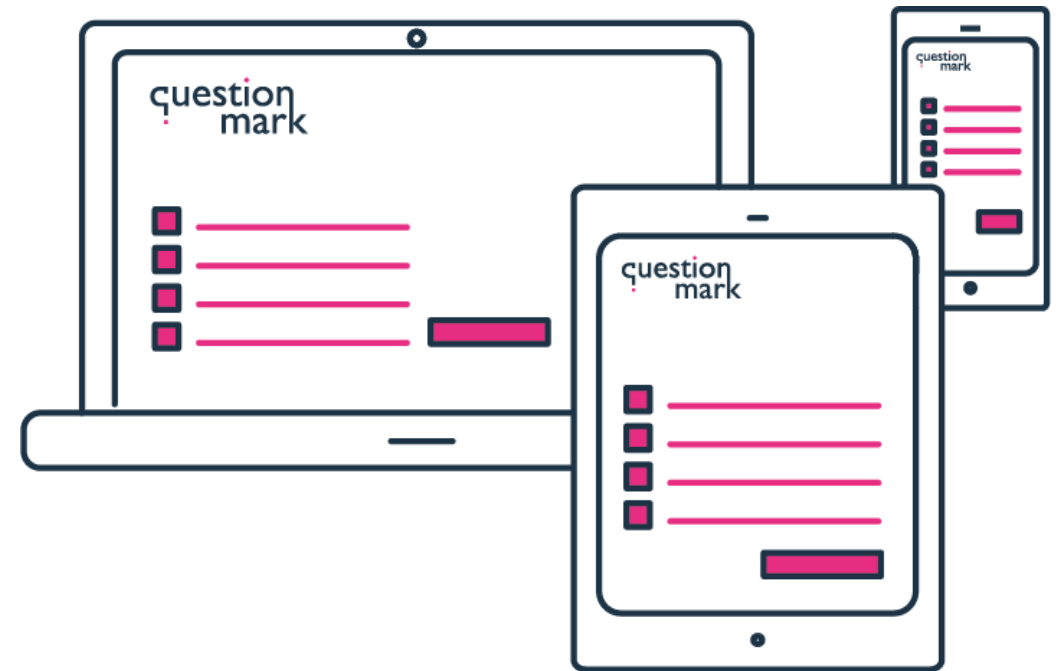
# Repeat Poll

## Which type of standard is better?

- Criterion referenced standard
- Norm referenced standard
- Don't know about either
- Neither is better/it depends
- Need a fusion of both

# Recap of agenda

- Introduce Norm Reference and Criterion Reference through examples
- Describe what each hopes to achieve
- Contrast both to learn about the differences
- Discuss standard setting, and standard maintenance for both
- Conclude and questions

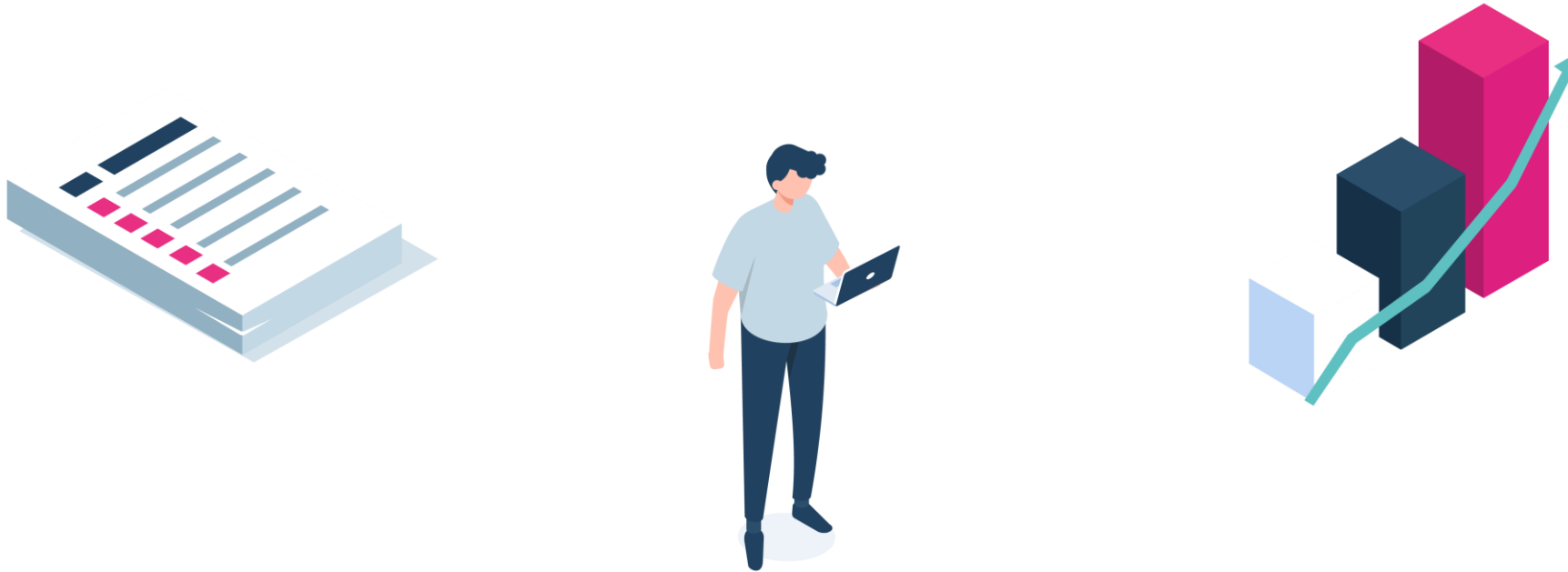


question  
mark

Questions?



# White papers, infographics, reports, eBooks and more!



**[www.questionmark.com/resources](http://www.questionmark.com/resources)**

# Upcoming webinars

## Introduction to Questionmark's Assessment Platform

- ◆ September 28, 2021 - 12:00 pm to 1:00 pm (EDT)
- ◆ October 12, 2021 - 12:00 pm to 1:00 pm (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](#)

---

## 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.

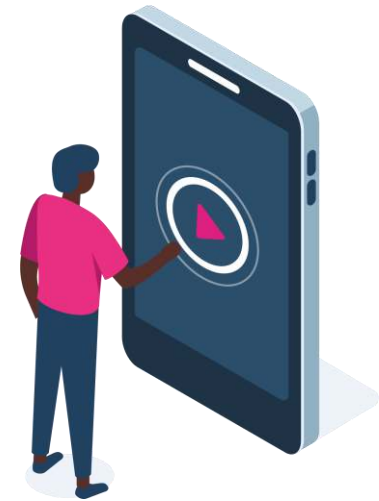
[Click to Register](#)

---

## Justifying the Value of a Certified Business Partner

- ◆ November 10, 2021 - 11:00 am to 12:00 pm (EDT)

Vendors depend on their Business Partners to reach customers, often in different segments, and provide solutions based on the vendor's key products. This webinar will cover why certifying your Business Partners is a business-smart decision.

[Click to Register](#)

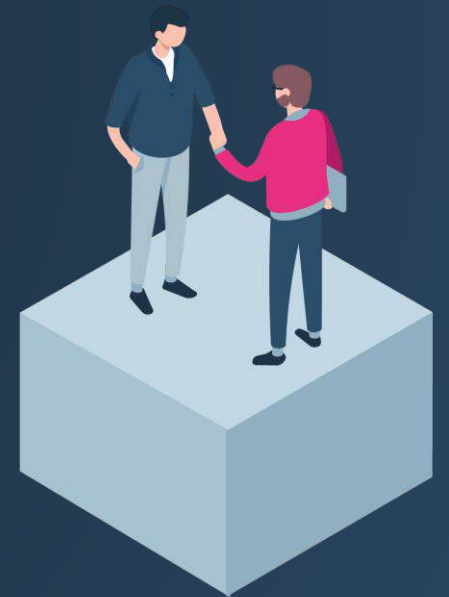


# How to Evaluate

## **Request a one-on-one demo**

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

[www.questionmark.com/request-demo](http://www.questionmark.com/request-demo)







# Thank you for attending!

***GET IN TOUCH WITH US***

*Reach out to Tom Gallacher – [Tom.Gallacher@Questionmark.com](mailto:Tom.Gallacher@Questionmark.com)*