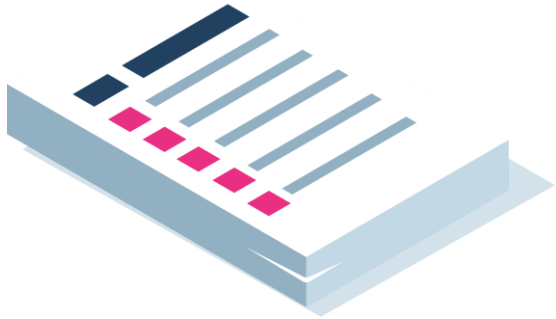
question mark

— a Learnosity company —

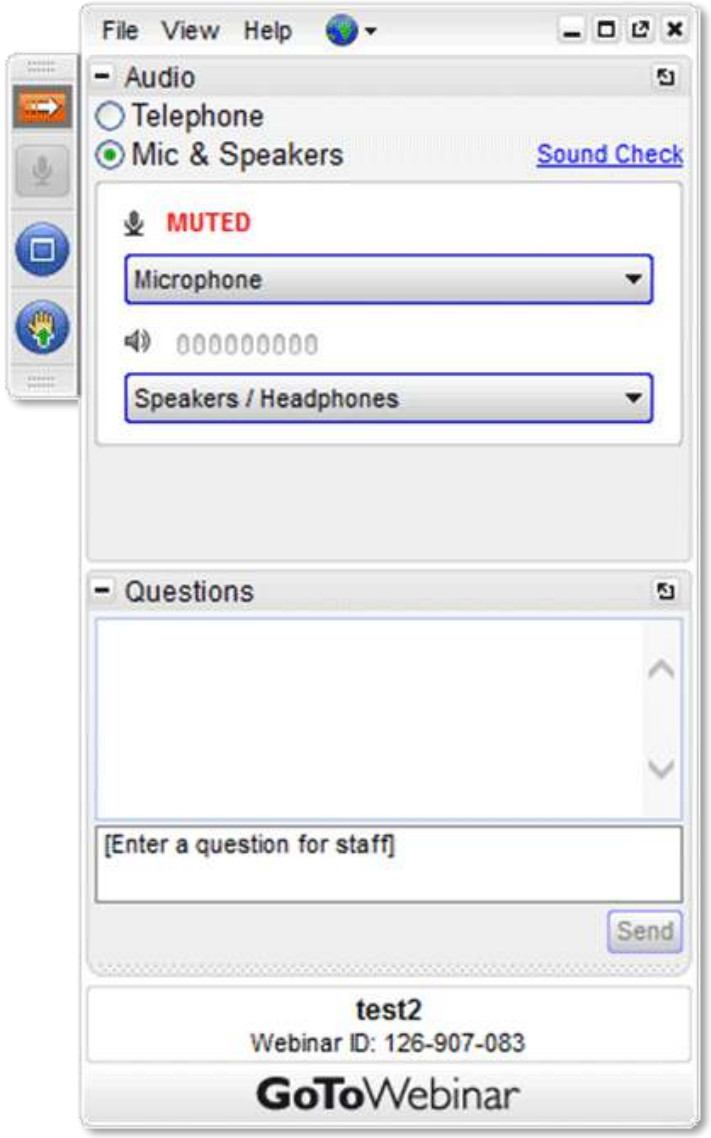


Make the Right Choice: How Many Multiple Choice Options is Optimal?

Tom Gallacher

Psychometrics Product Manager

To ask questions, use the “Questions” feature



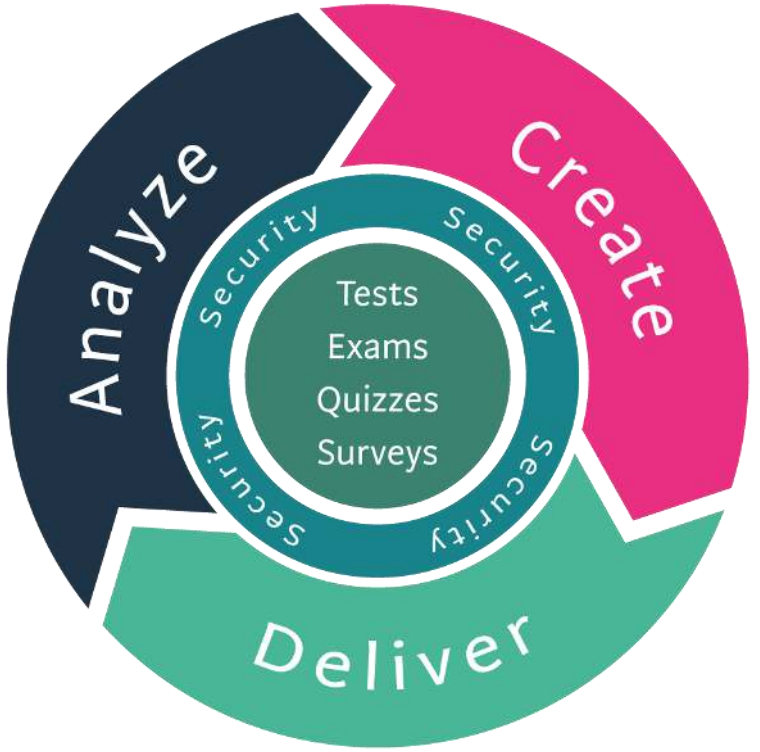
Watch for an email after the webinar:

- Download slides (PDF)
- View a recording

About Questionmark

Background

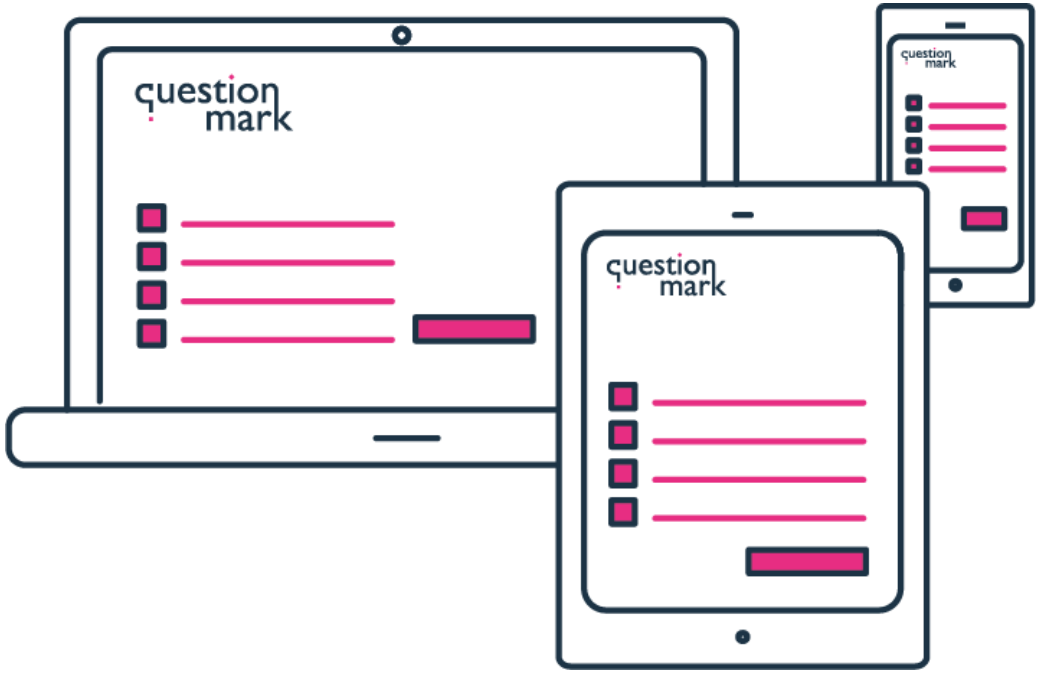
- Founded in 1988
- Assessment solutions to measure knowledge, skills, abilities and attitudes securely for certification, regulatory compliance, workforce learning, sales-force readiness and higher education
- ISO/IEC 27001 Certified (Learn more: www.questionmark.com/trust)



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

Agenda

- Discuss practicalities of writing good MCQs
- Introduce concept of ‘test information’
- Propose an alternative to ‘negative marking’
- Present findings from a meta-analysis
- Provide actionable tips for how to implement the findings



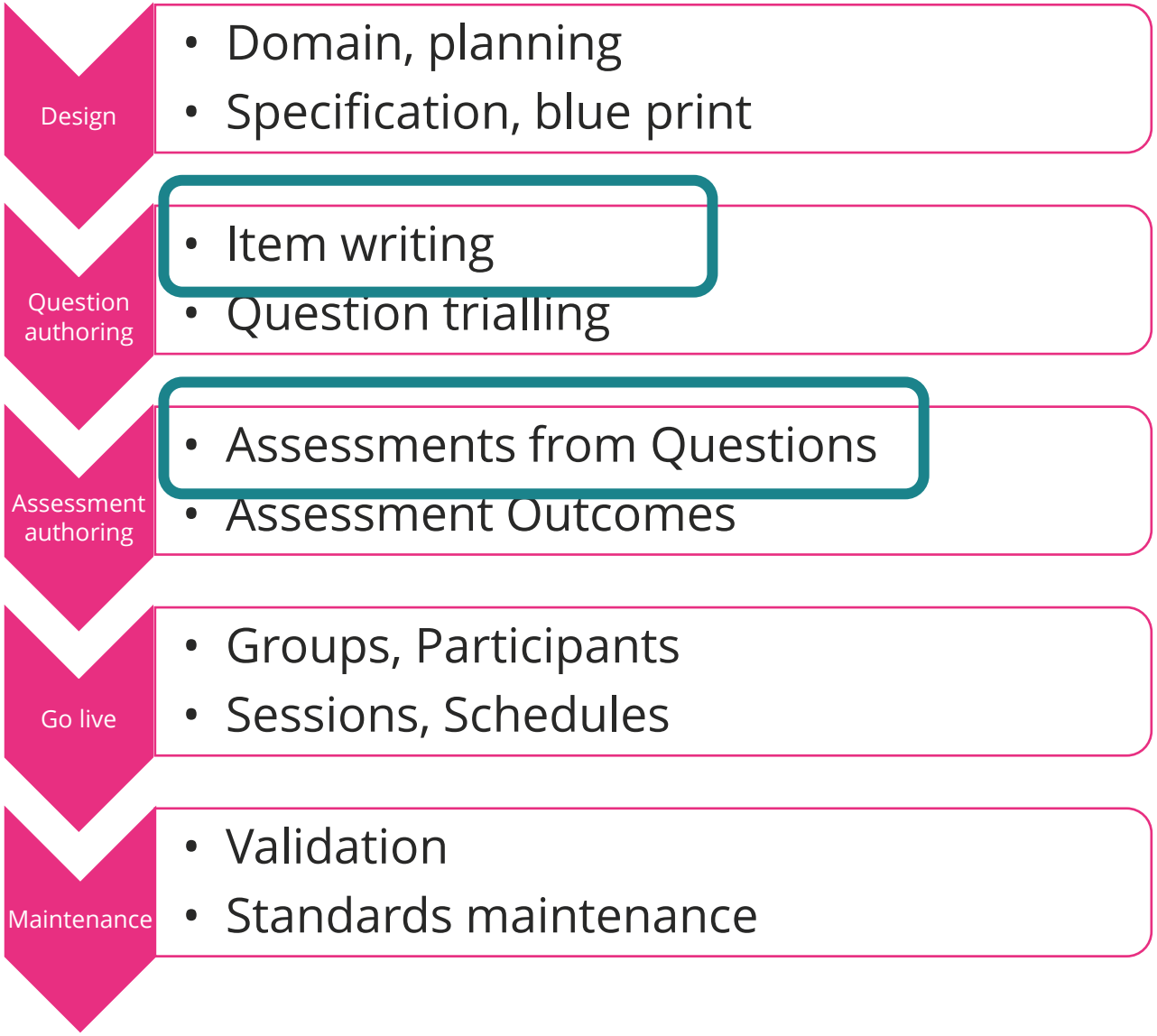
About the presenter

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.

High level view

- Design: planning the assessment and working backwards
- End up with high quality assessment, decisions are valid, estimates are reliable, participants are motivated
- There are lots of materials to help with each stage
- This webinar will deal with some finer points of item writing for a commonly used question type



Some general principles of item writing

Do's

- Write with the participants in mind: what they know, what they might not
- Express the demands of the question clearly: no trick questions.
- Give the resources that the participant needs to answer the question (think about scrap paper and pen?)
- Map the question demands back to a part of the domain
- Cover the domain adequately
- Keep track of 'friends and foes' or 'parents and clones'

Don'ts

- Do not include in questions things that can be distressing, or culturally sensitive
  - Dogs as pets, spiders, snakes, insult to the deceased, insult to the elderly, extra-terrestrial life, consumption of alcohol…
- Avoid questions that can be answered with knowledge outside the domain
  - E.g. sports, music trivia

Multiple Choice Questions

- Task is to identify which of competing options answers the question the best
- Ideally, question is such that there is only one correct answer
- Reading options takes participants' time
- Writing options takes authors' time
- More options, less chance of randomly guessing correctly

What is the capital city of Nigeria? Choose one of the following.

A. Lagos

B. Abuja 

C. Accra

D. Benin City

E. Niamey

F. Malabo

question mark
— a Learnosity company —

Quick Poll 

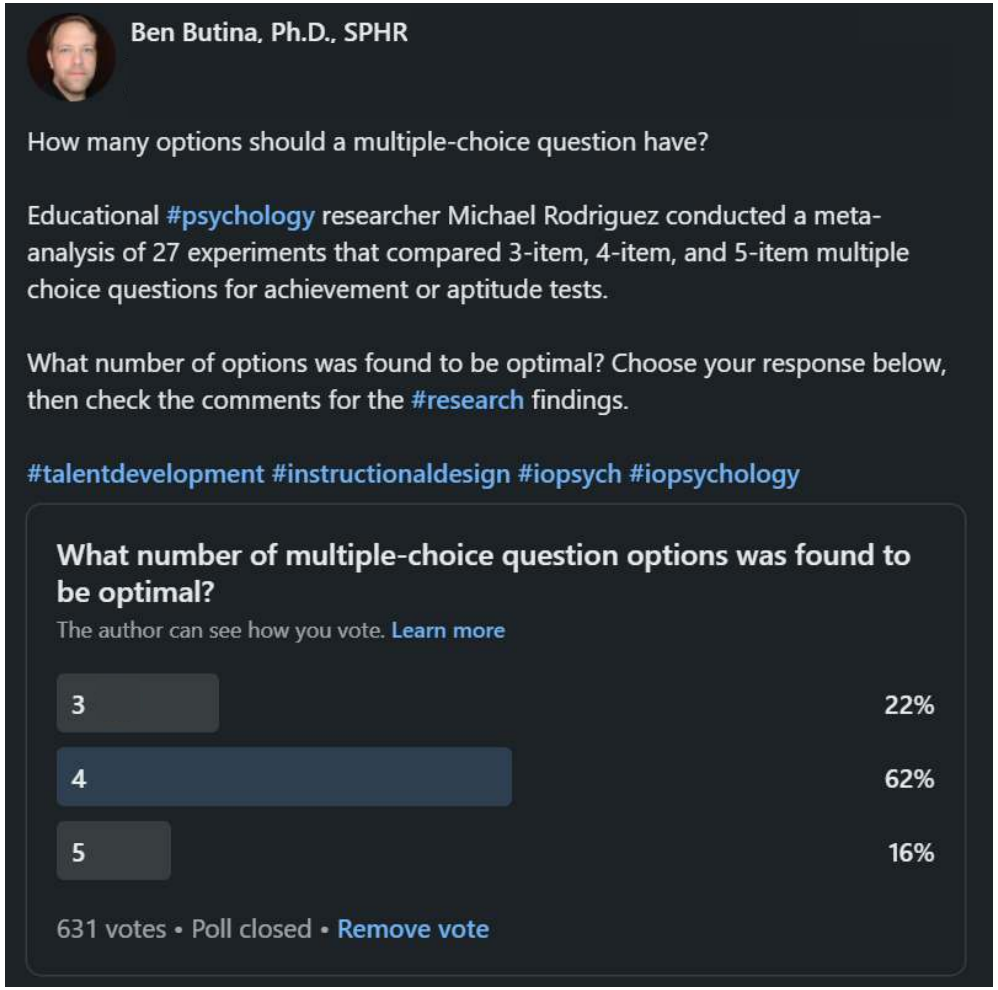
How many MCQ options are optimal?

- 3 or fewer responses are optimal
- 4 responses are optimal
- More than 4 responses are optimal

Optimal for what?

- Quality → validity → reliability → information
- "Test information" is a concept in Item Response Theory
  - We don't need the formula right now, just the concept
  - Test information is a function of item information
  - More questions means more test information
  - Item information is maximized when participants are given questions that they are 50% likely to get correct
  - Why we typically aim for questions that have mean scores of 50%
- Particularly in the practical scenario where time is limited:
  - We want the most test information to inform estimates as possible
  - How many options gives the most item information per time
- Optimal tests maximize item information & test information

My prior distribution



- My “prior distribution” was informed by this poll on LinkedIn.com
- Four options is pervasive
- Where does four come from?
- Qualitative review of the comments suggested that specifications don’t change very much
- But a trade off between costs and test information?

question mark

— a Learnosity company —

Guessing

The bad guy, or just misunderstood?

Guessing correctly weakens correlations

- Question-total correlation, or discrimination index are both affected by who gets the correct answer.
- Metrics of question quality.
- When low trait participants have a high likelihood of guessing correctly, then we typically throw the question out
- Because getting the question correct doesn't provide information on the ability.

Plug for the Item Analysis Dashboard!

- In the 2021-Q4 release we put out an Item Analysis Dashboard that will come to replicate and extend some of our legacy reports.
- Feedback welcome.

Randomly guessing correctly?

- Sometimes it is useful to consider the score achieved if participants were to guess randomly

- If there are results below this then:
  - Are these participants really trying to maximize their scores,
  - Are they disengaged,
  - Are they misunderstanding the assessment demands?

- But often, participants guess *intelligently*
  - I.e. “I know it’s not C, it could be A or B so lets choose A”.

Penalizing guessing

- Strategy commonly known as ‘negative marking’
- If guessing tends to be intelligent rather than random, two participants of different abilities will correctly guess a different number of questions
- So even guessed responses can be used to inform ability estimates
- But item statistics will be expected to take a hit
- Penalising guessing might improve the item statistics, while removing test information
- So test might look more reliable in some stats, but less reliable when it comes to the important bit: estimating performance.

Penalizing guessing; systematic effects?

- Evidence suggests that male participants are more likely to guess when they don't know the answer.
- Risk averse students are less likely to guess.
- Is the point of assessment to cause anxiety?

- What this means for tests with negative marking, are we penalising female participants?

Baldgia, K. (2014). Gender differences in willingness to guess. *Management Science, 60*, pp.434:448. DOI: 10.1287/mnsc.2013.1776

Coffman, K.B., & Klinowski, D. (2020). The impact of penalties for wrong answers on the gender gap in test scores. *PNAS*, *117,* pp.8794-8803. DOI: 10.1073/pnas.1920945117

‘Up the hill on low’

- Isn’t the timer trying to cause anxiety?
- Test performance “is like ‘power’ as the physicist uses the word: the amount of work that can be done in a given time…
- “A frequent complaint made of the tests is that they place too much emphasis upon speed…
- “If… people have less power, they have to go up the hill on low gear and it takes them longer; that is all. Of course they ‘get there’ just the same, but when they ‘get there’ their powerful rivals are on and somewhere else”

Boring, E. (1923). Intelligence as the Tests Test It. *New Republic, 36,* pp.35-37

By Nicolas Langlois-Demers, CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Cycling_in_Fraser%27s_Hill_July_2017.png

Rewarding correct elimination?

- An alternative assessment technique is to not have MCQ at all, rather repurpose the same item content, but with different marking
- Rather than 1 point for 1 correct response identified, have n-1 points awarded for eliminating incorrect responses.
- This type of question is possible with Questionmark's Multiple Response.
- Minimal test gender bias, increased participant performance, satisfaction, reducing anxiety.
- Is definitely an interesting area worthy of further research if anybody is interested in collaboration?

Bond, A.E., Bodger, O., Skibinski, D.O.F., Jones, D.H., Restall, C.J., Dudley, E., van Keulen, G. (2013). Negatively-marked MCQ assessments that reward partial knowledge do not introduce gender bias yet increase student performance and satisfaction and reduce anxiety. *PLoS ONE, 8,* e55956. DOI:10.1371/journal.pone.0055956

question mark

— a Learnosity company —

MCQ options

The optimal number?

So more distractors is better?

- Say we're committed to MCQ.
- If we have more distractors, less chance of randomly guessing correct
- But that costs us, and it costs the participant time at least.
- Often a struggle to write good distractors
- Often, some distractors are so implausible that we find that no participant select them
- Diminishing returns for each new MCQ option written.

What is the capital city of Nigeria? Choose one of the following.

A. Lagos

B. Abuja 

C. Accra

D. Benin City

E. Niamey

F. Malabo

G. Ogbomosho

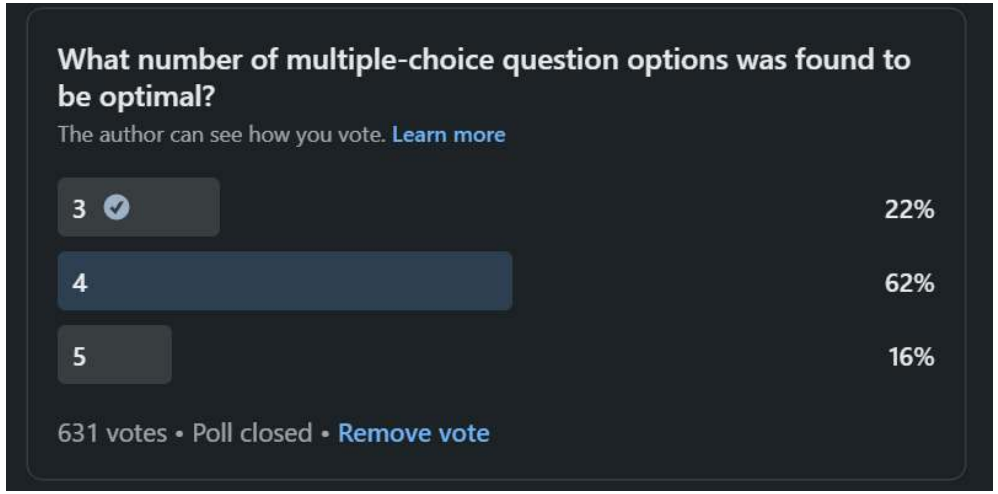
H. Ibadan

3 is a magic number

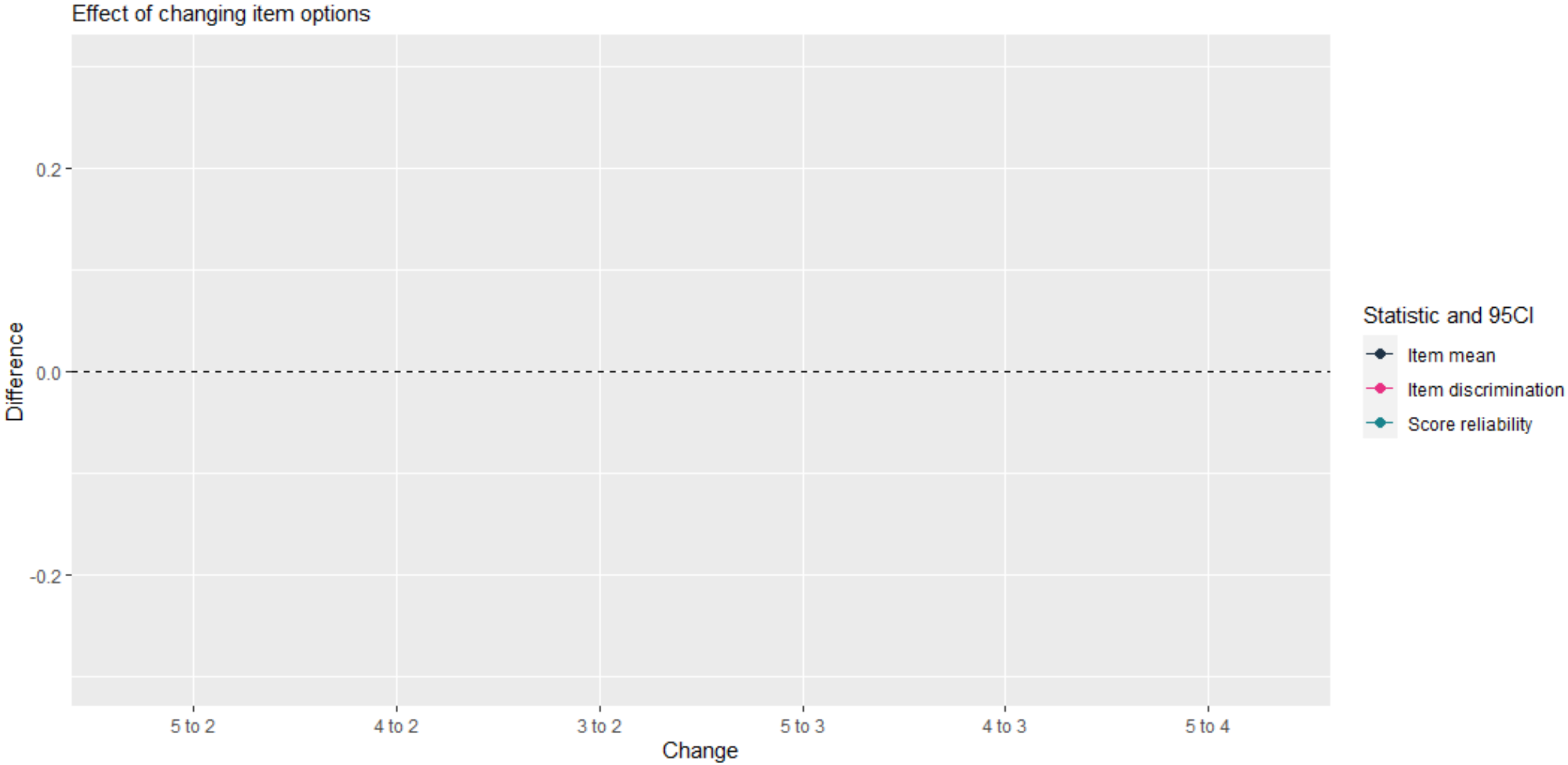
- **Current meta-analysis says that 3 is the optimal number of options**
- This has been argued in 1944, 1964 and 1977.

Rodriguez, M.C. (2005). Three options are optimal for Multiple-Choice Items: A meta-analysis of 80 years of research. *Education Measurement: Issues and Practice, 24*, pp.3-13, DOI:10.1111/j.1745-3992.2005.00006.x

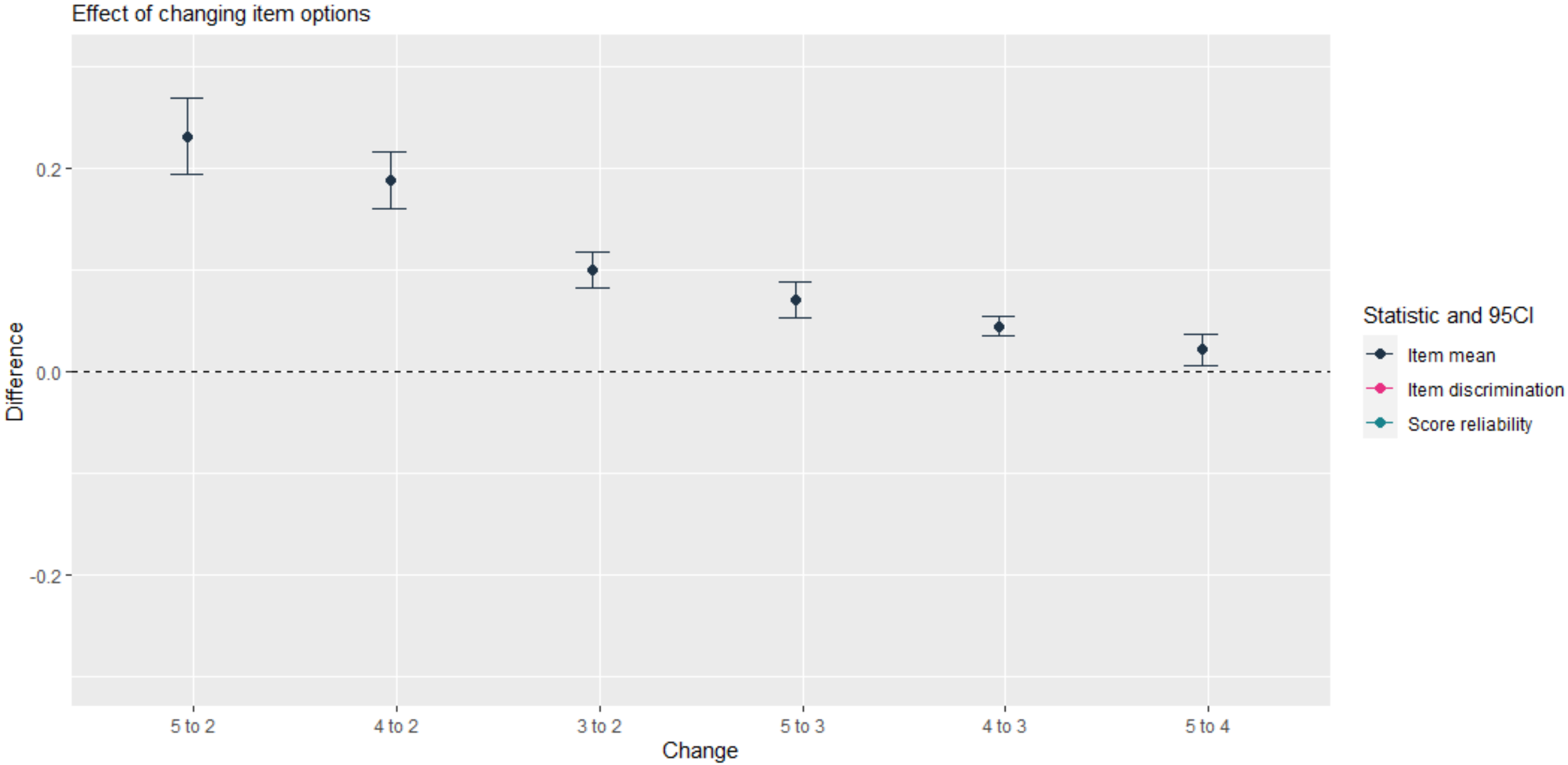
- "Whenever the amount of time spent on the test is proportional to its total number of alternatives, the use of three alternatives at each choice point will maximize the amount of information obtained per time unit" – Tversky 1964



Make the right choice

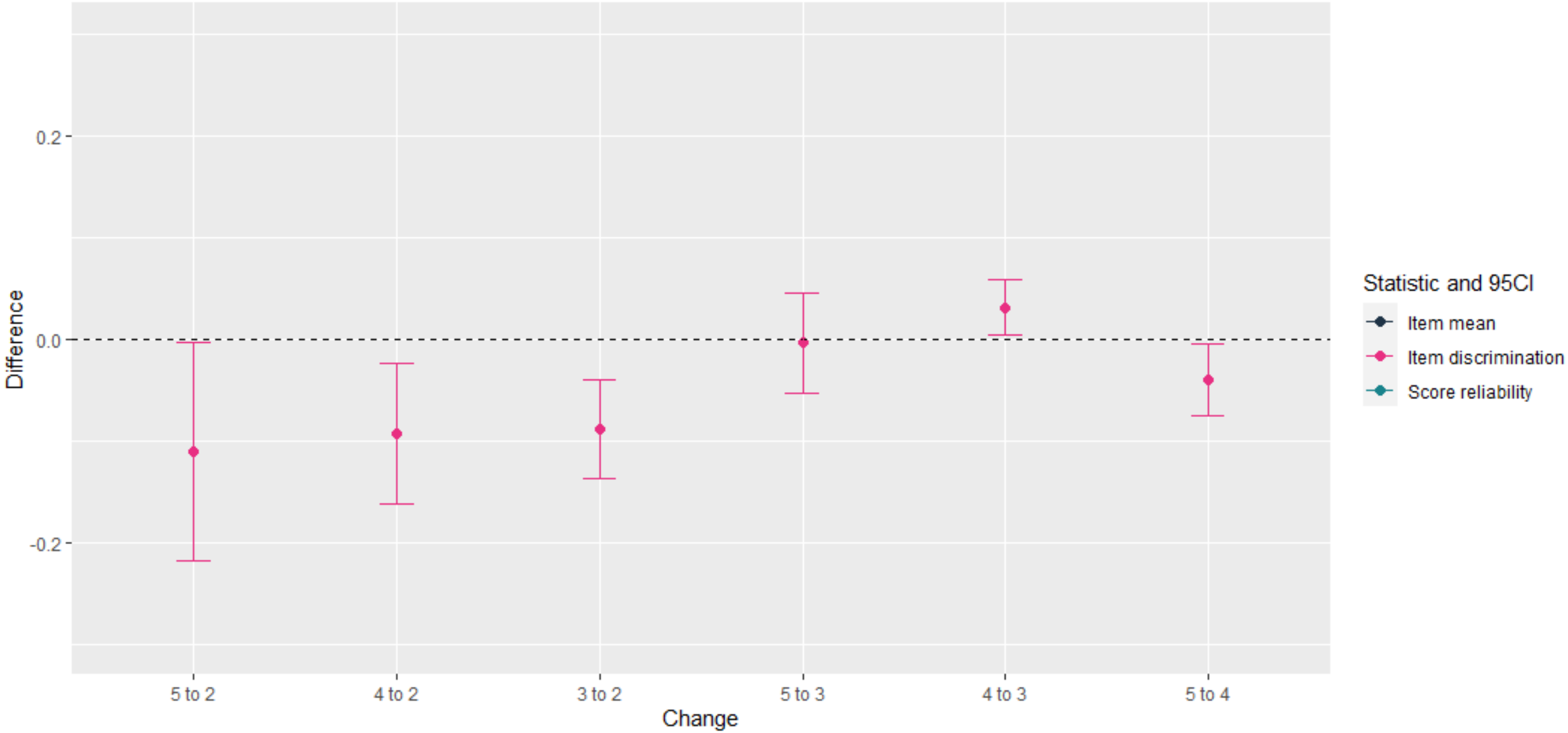


Make the right choice

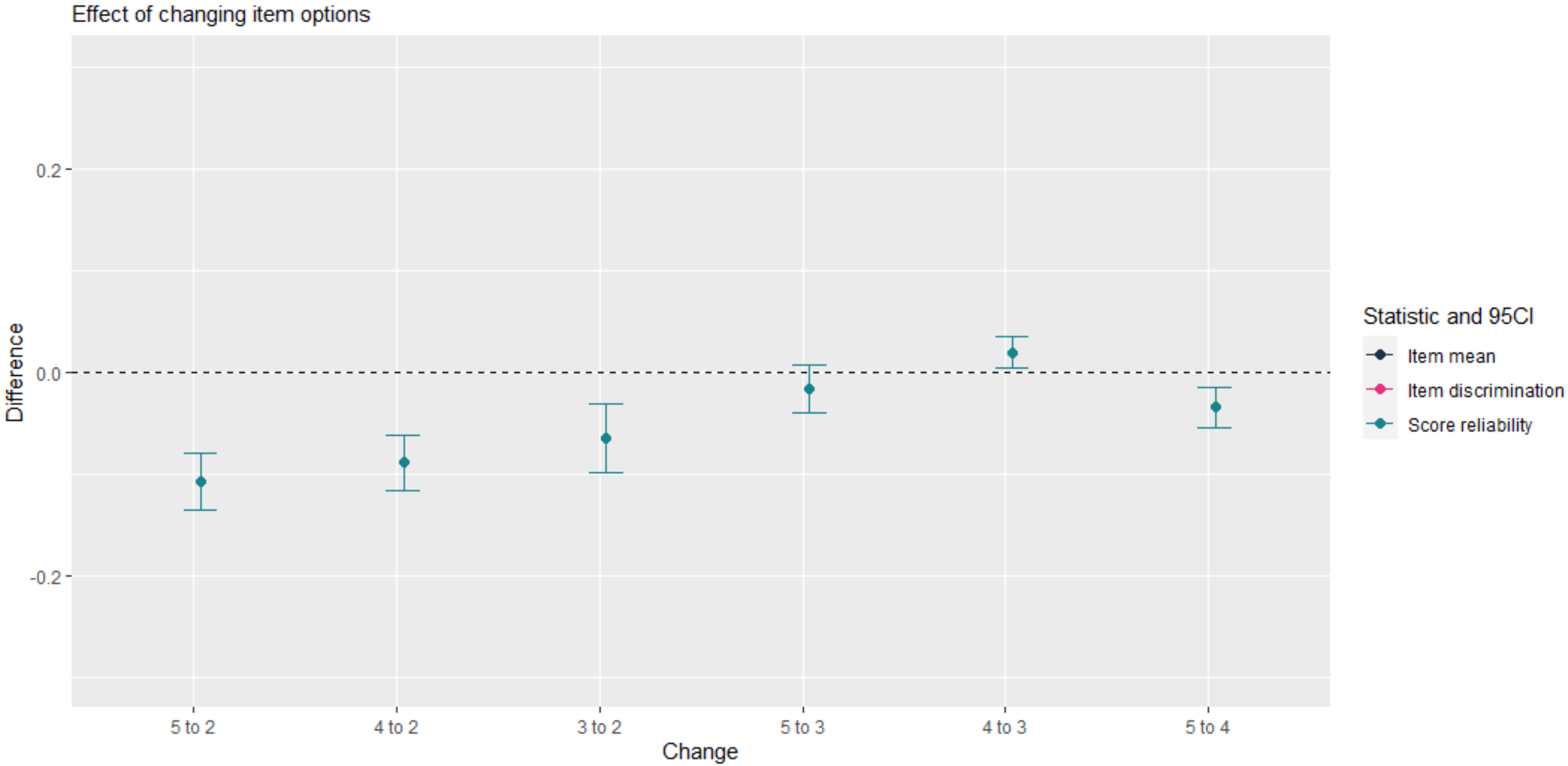


Make the right choice

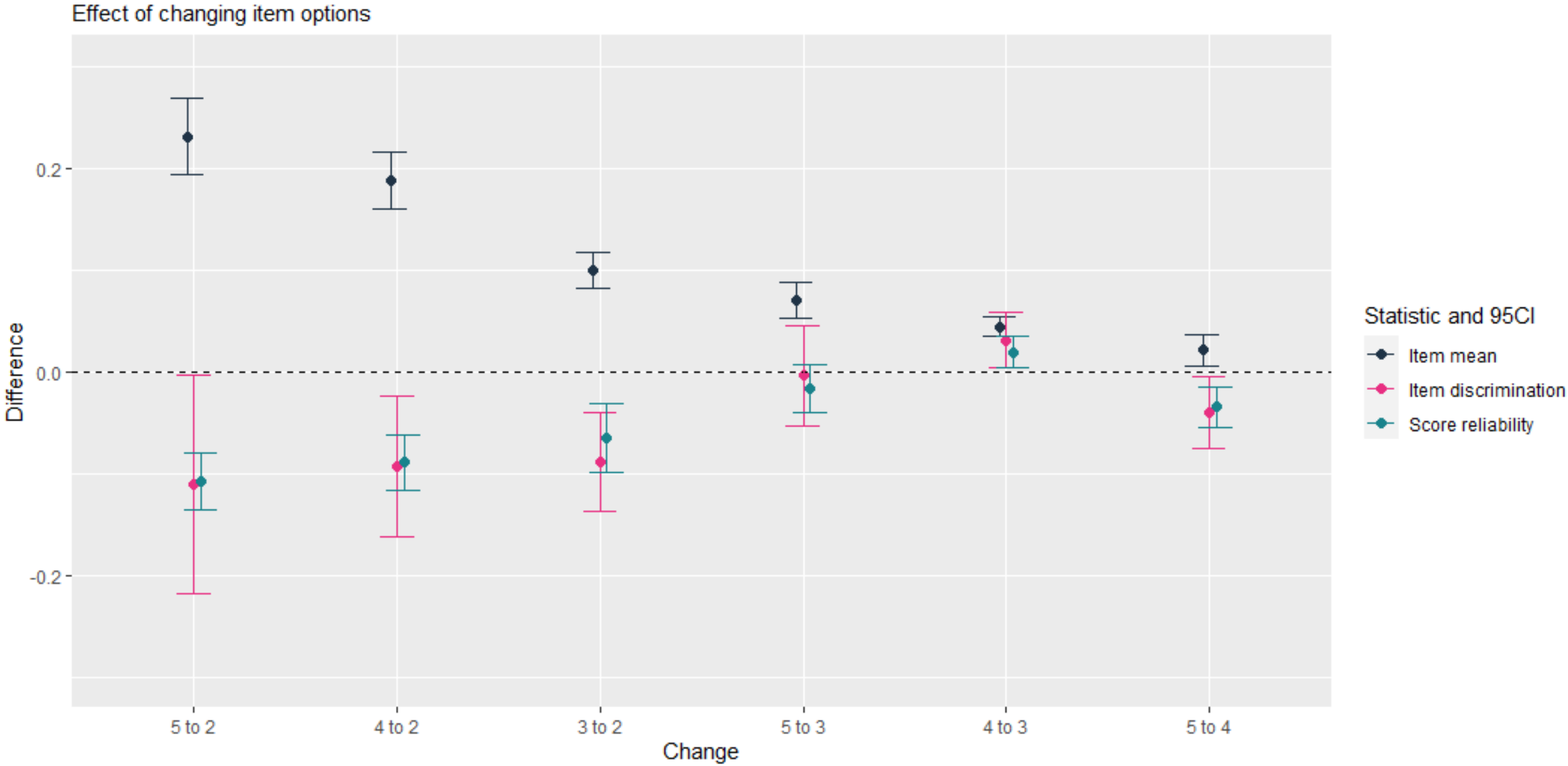
Effect of changing item options



Make the right choice



Make the right choice



Impacts of reducing options

- Higher question mean (easier)
- Negligibly lower question-total correlation (less reliable)
- Negligibly lower reliability (less reliable)
- But going down to three seems to buck the last two.

- Particularly if we're reducing by removing the least chosen distractor.
- Allows adding in more questions without affecting time!
  - Improves reliability, information, which might improve validity and quality

question mark

— a Learnosity company —

Conclusions and recommendations

Some practical tips

Practical tips

- If your assessments use a timer:
- Author multiple choice questions with 4 distractors
- Trial the questions to know which distractor is the least chosen
- Use the new Item Analysis dashboard in Questionmark
- Drop the least chosen option and republish the questions
- Have more questions in the assessment

- Remember that the timer is not to stress out participants, it is to identify ‘power’: mental output in a fixed time.
- But that those who might have less ‘power’ can still get up the hill.

Practical tips II

- If timer is not used:
- Not invalid to have >3, but it is worth considering
- If we don't want to change how many items are administered, then under some circumstances we might be (marginally) better off with five than with the same number of four option or three option questions.
- If we need to make questions harder then we can add more options but it is inefficient.

- While the research has looked at fixed form linear testing, it would be interesting to hypothesize about effects on Computer Adaptive Testing.

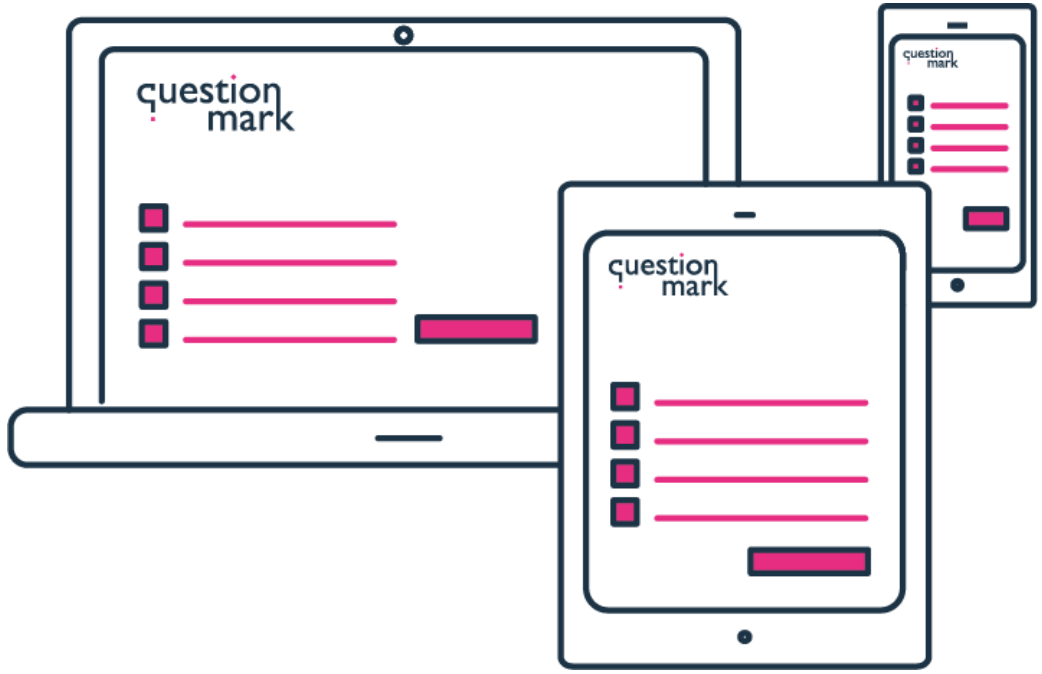
Negative marking

- Doesn’t sit well with me
- Why not try rewarding elimination instead?

- Nothing in assessment should cause any more anxiety than necessary
- Try to maximize the participants’ performance
- Improving the item statistics at the cost of information may improve reliability but wont improve validity

Agenda

- Discuss practicalities of writing good MCQs
- Introduce concept of ‘test information’
- Propose an alternative to ‘negative marking’
- Present findings from a meta-analysis
- Propose actionable tips for how to implement the findings

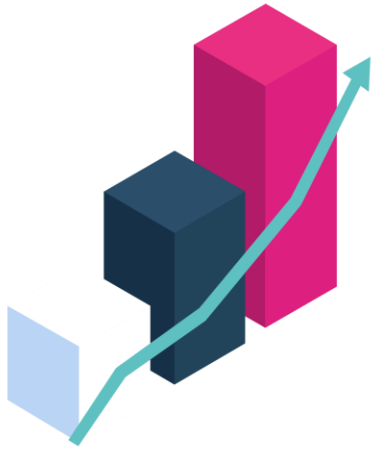


question mark

— a Learnosity company —

Questions?

White papers, infographics, reports, eBooks and more!



www.questionmark.com/resources

Upcoming webinars

Introduction to Questionmark's Assessment Platform

- February 22, 2022 - 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.

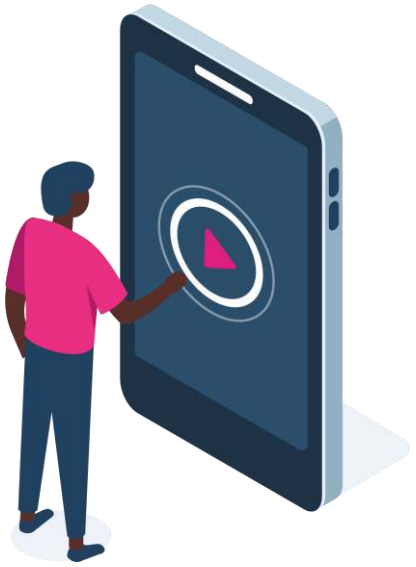
Click to Register

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

- March 8, 2022 - 11:00 am to 12:00 pm (EDT)

Confused about the difference between validity and reliability? Looking for ideas on what goes into to putting together a good test? This session will explore strategies for developing assessments that are both valid and reliable to provide stakeholders with actionable, defensible results for informed decision-making.

Click to Register



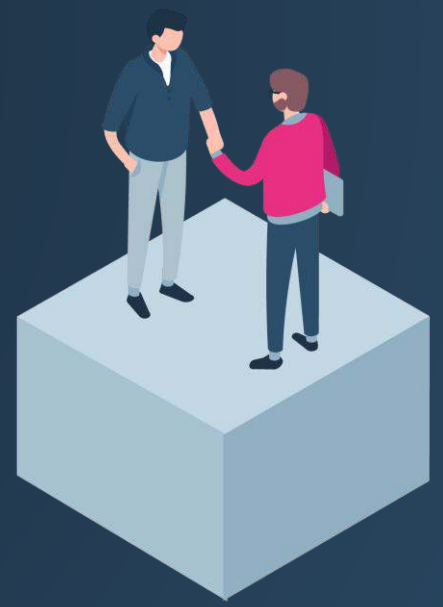
— a Learnosity company —

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



question mark

— a Learnosity company —

Thank you for attending!

GET IN TOUCH WITH US

Reach out to Tom Gallacher – Tom.Gallacher@Questionmark.com