

A person is sitting at a desk, writing in a notebook with a pen. There is a white mug on the desk to the left, and a tablet is visible in the foreground. The background is slightly blurred, showing another person in a blue shirt.

COLLABORATIVE TEST DEVELOPMENT

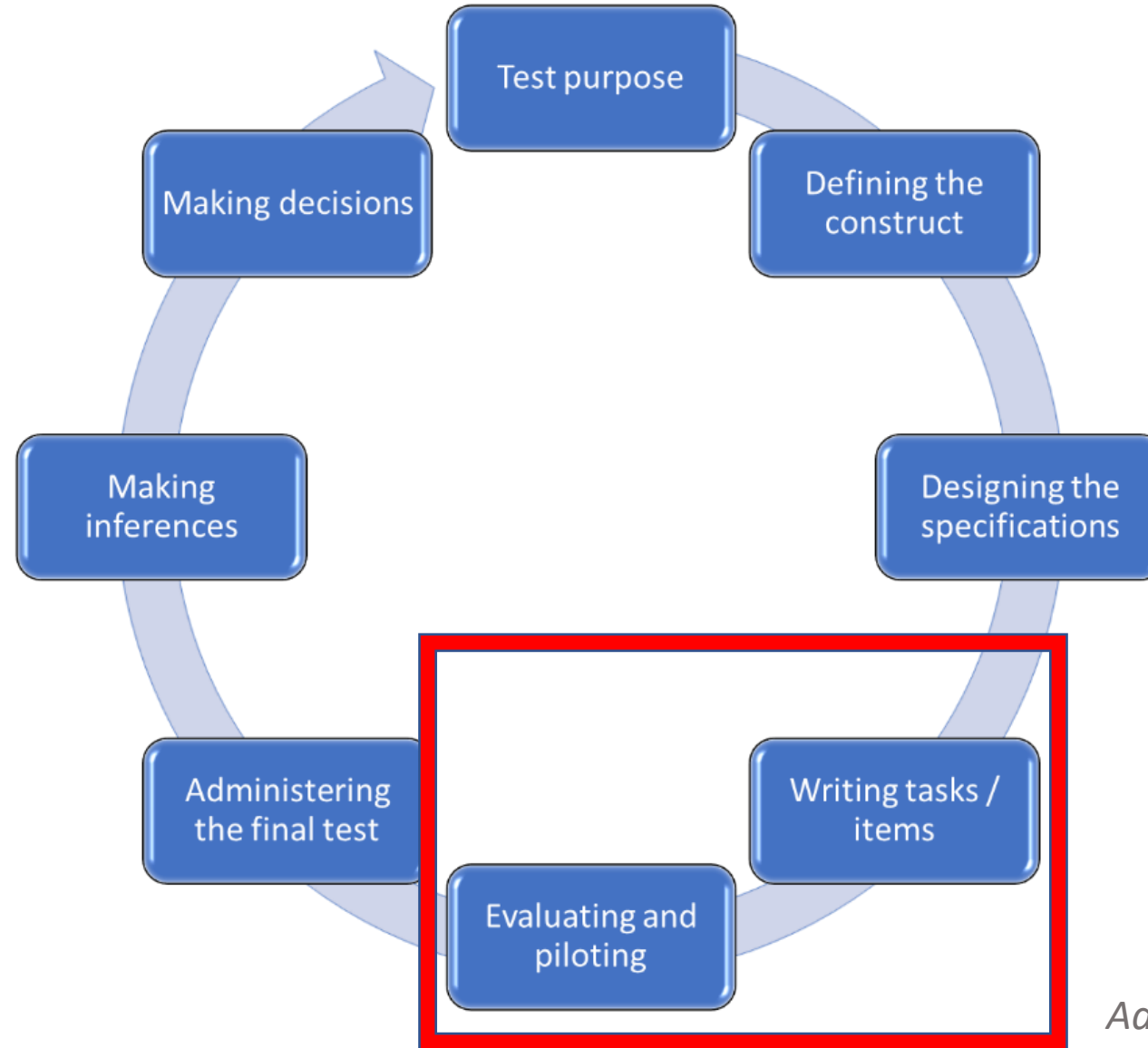
the role of item moderation in ensuring test quality

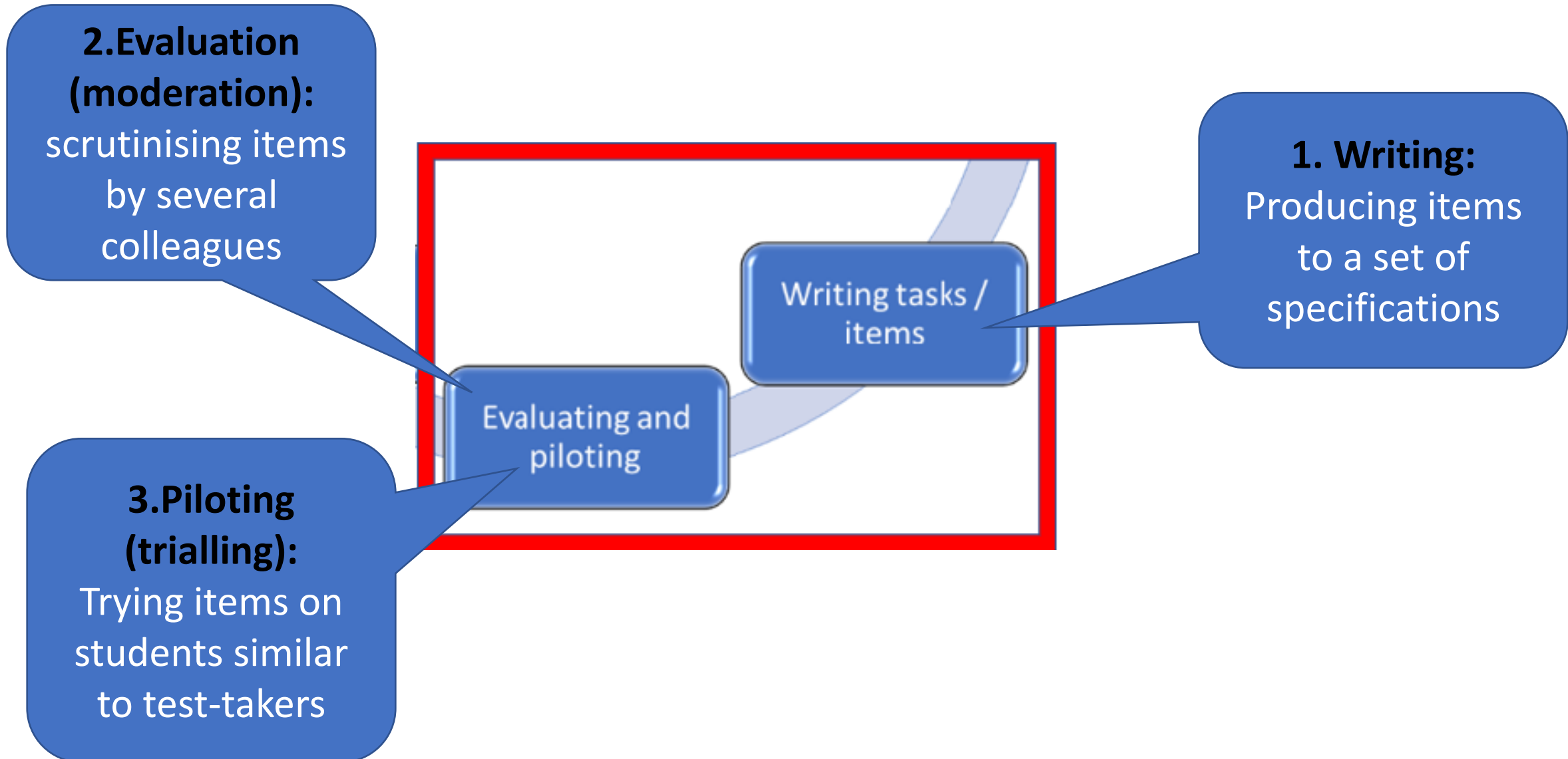
OLENA ROSSI, PhD

o.rossi@lancaster.ac.uk

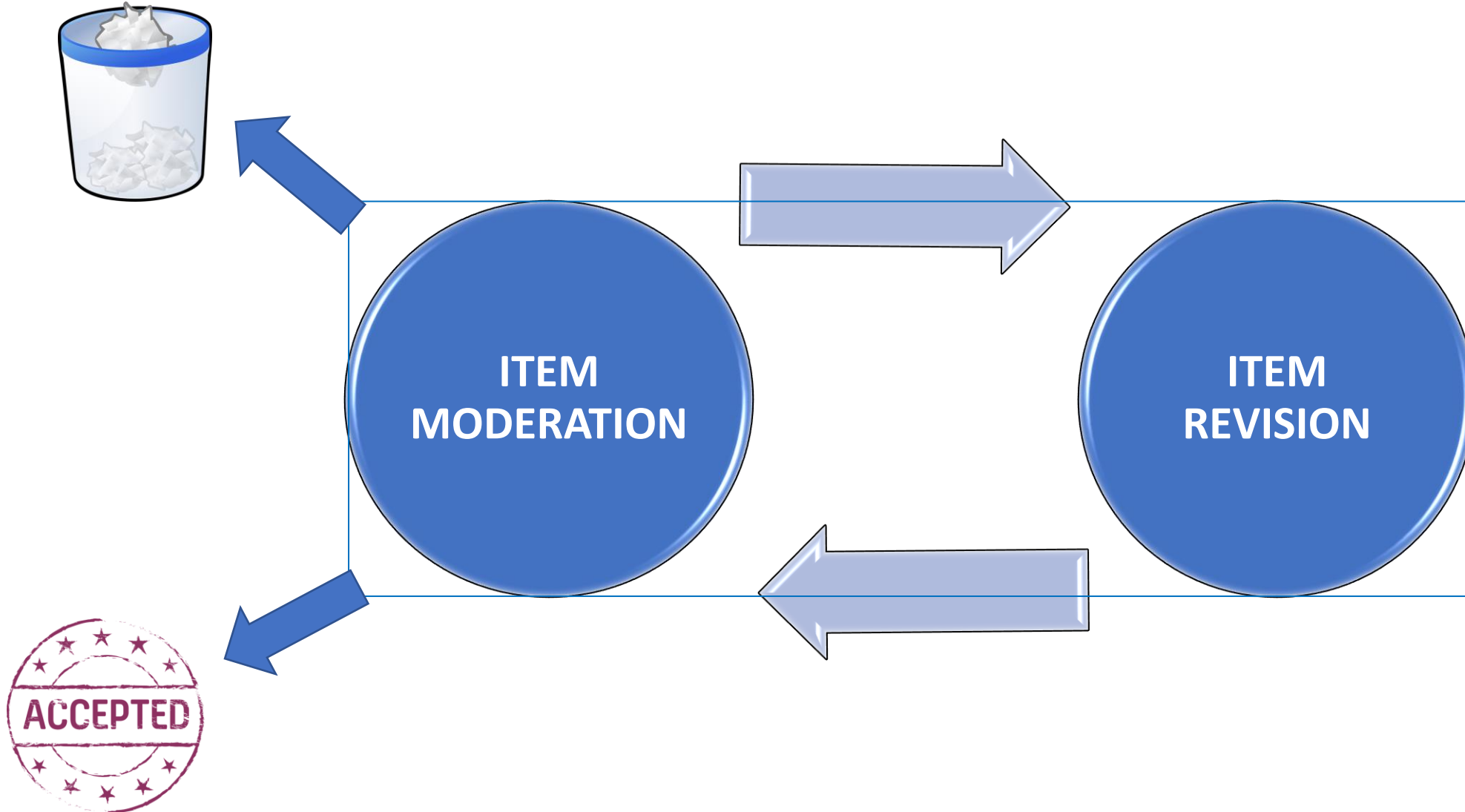
olena.rossi@itemwriting.co

Test design cycle





Iterative item moderation



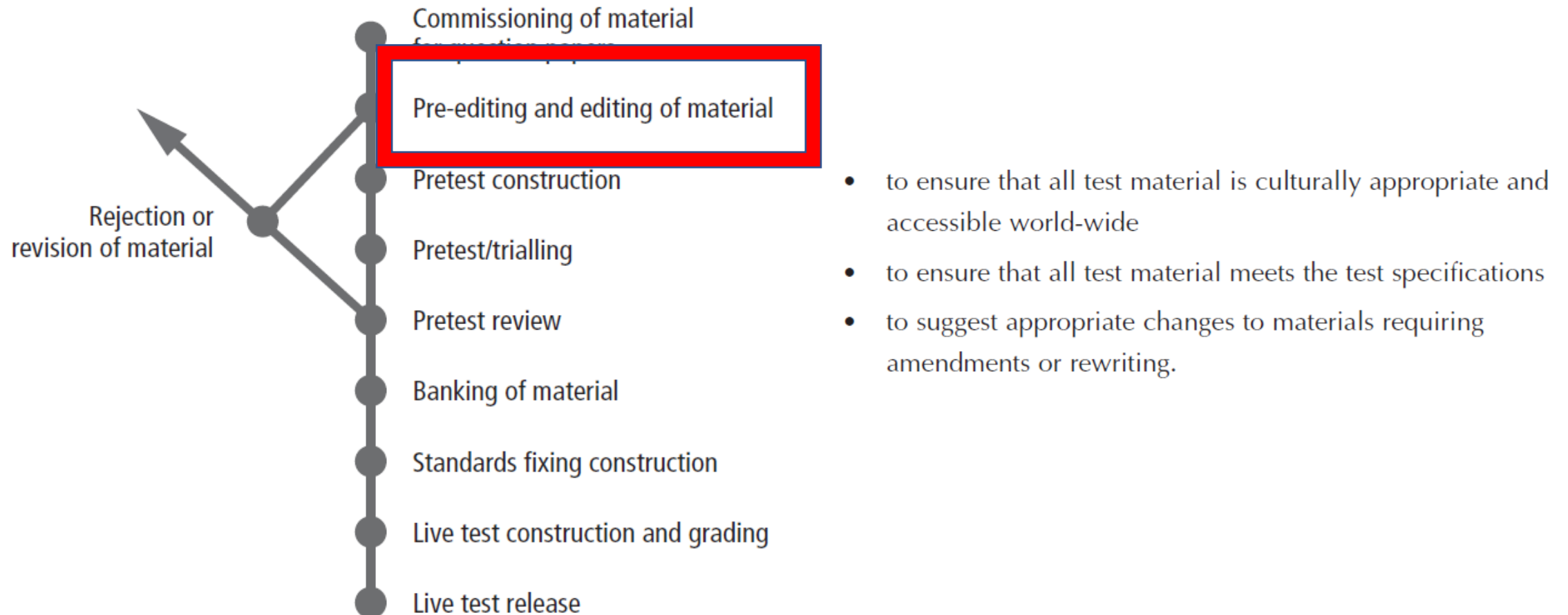
HOW DOES ITEM WRITING HAPPEN IN YOUR ORGANISATION?

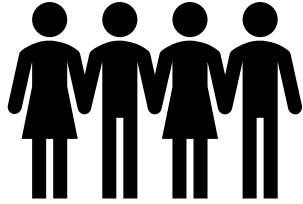
- What process do you follow?
- Do you have an item moderation stage?
If yes, how do you moderate items?



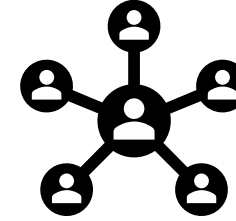
Item moderation at large-scale testing organisations

Figure 1: The Question Paper Production process at Cambridge ESOL

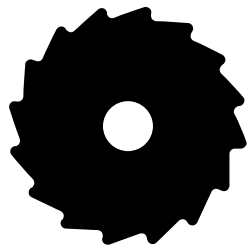




1. Set up an item moderation group: at least 3 item writers, from the same organisation or inter-organisational



2. Create a Google Docs folder every group member has access to



3. Establish item moderation routines: who moderates what, type of feedback, deadlines




4. Moderation meeting

B2 item – 102 Relative clauses	
Stem	There was a very well reviewed film _____ last week.
Option 1	which were seen
Option 2	that she sees
Option 3	that I saw *
Option 4	which I see

B2 item – 102 Relative clauses

First, do the item as a test-taker!

 Olena Rossi
4:52 PM Today ✓
This distractor seems to target tenses and not relative clauses.



 Olena Rossi
4:52 PM Today ✓
This distractor also targets tenses.

TABLE 2: MODERATION OF GRAMMAR ITEMS

1. Is the English grammatically correct?
2. Is the English natural and acceptable?
3. Is the English in accordance with the specifications?
4. Does the item test what it is supposed to test, as specified?
5. The correct response cannot be obtained without the appropriate knowledge of grammar (other than by random guessing)
6. Is the item economical?
7. a. Multiple choice - is there just one correct response? b. Gap filling - are there just one or two correct responses?
8. Multiple choice: Are all the distractors likely to distract?
9. Is the key complete and correct?

Hughes & Hughes, 2020, p.68

CHECKLIST		Yes	No
 Reviewing Multiple-Choice Items			
1. Is this the most appropriate type of item to use?		___	___
2. Does each item stem present a meaningful problem?		___	___
3. Are the item stems free of irrelevant material?		___	___
4. Are the item stems stated in positive terms (if possible)?		___	___
5. If used, has negative wording been given special emphasis (e.g., capitalized)?		___	___
6. Are the alternatives grammatically consistent with the item stem?		___	___
7. Are the alternative answers brief and free of unnecessary words?		___	___
8. Are the alternatives similar in length and form?		___	___
9. Is there only one correct or clearly best answer?		___	___
10. Are the distracters plausible to low achievers?		___	___
11. Are the items free of verbal clues to the answer?		___	___
12. Are verbal alternatives in alphabetical order?		___	___
13. Are numerical alternatives in numerical order?		___	___
14. Have <i>none of the above</i> and <i>all of the above</i> been avoided (or used sparingly and appropriately)?		___	___
15. If revised, are the items still relevant to the intended learning outcomes?		___	___
16. Have the items been set aside for a time before reviewing them?		___	___

Miller et al., 2009, p.215

Item moderation checklists

Multiple Choice

Moderation criterion	YES	NO	What exactly is the problem? How can the item be improved?
Item stem			
Concise			
Clear			
NOT worded negatively (e.g. "What did Anna <u>not</u> do?")			
Options			
Concise			
Clear			
There are no repeating words that can be moved to the stem			
Similar in length			
Similar in form/content			
NOT "none of the above", "all of the above"			
Key			
Looks similar to the distractors (does not stand out)			

Space for explanations and suggestions for improvement

Where do the criteria come from?

1. Item specifications
2. Previous moderation experience
3. Best practice for producing items of this type

1. Item specifications

MULTIPLE CHOICE GRAMMAR ITEM SPECIFICATIONS

For a general English language proficiency test, international adult candidates

Specifications	B2
Task description	Sentence completion based on the appropriacy of grammatical meaning and/or form
Format	4-option multiple choice
# items	1
Word count - stem	Max. 15 words
Word count - options	1-3 words
Key	Indicate with *
Lexical level	K1-K4
Grammar (exponent)	B2: 102 Relative clauses
Topic	Appropriate at B2 level
Function	Appropriate at B2 level

The item is at B2 level

The item targets the meaning or form of the grammar structure

The intended construct is targeted

2. Previous moderation experience

MULTIPLE CHOICE GRAMMAR ITEM SPECIFICATIONS	
For a general English language proficiency test, international adult candidates	
Specifications	B2
Task description	Sentence completion based on the appropriacy of grammatical meaning and/or form
Format	4-option multiple choice
# items	1
Word count - stem	Max. 15 words
Word count - options	1-3 words
Key	Indicate with *
Lexical level	K1-K4
Grammar (exponent)	B2: 102 Relative clauses
Topic	Appropriate at B2 level
Function	Appropriate at B2 level

The stem provides enough context to ensure that the intended construct is tested

3. Best item-writing practice

MULTIPLE-CHOICE ITEMS

SHORT-ANSWER QUESTIONS

Moderation criterion	YES	NO	What exactly is the problem? How can the item be improved?
Item stem			
Concise			
Clear			
NOT worded negatively (e.g. "What did Anna <u>not</u> do?")			
Options			
Concise			
Clear			
There are no repeating words that can be moved to the stem			
Similar in length			
Similar in form/content			
NOT "none of the above", "all of the above"			
Key			
Looks similar to the distractors (does not stand out)			
Is truly correct as the answer			
Distractors			
Plausible			
NOT ridiculous or humorous			
Cannot serve as the correct answer			

Moderation criterion	YES	NO	What exactly is the problem? How can the item be improved?
Instructions			
State the format of the expected response (should it be a number, word, phrase, sentence?)			
State the response word limit			
State whether the response should come verbatim from the text or whether paraphrases are accepted			
Example			
The first item serves as an example			
The example complies with the instructions			
Questions			
Concise and clear			
Specific and precise – written in such a way as to limit the range of acceptable responses			
Open-ended			
Answer key			
As much as possible, includes all acceptable variations of the correct answer			
Does not require correct spelling (unless spelling is part of the construct)			
Does not require correct grammar (unless grammar is part of the			



itemwriting.co

Item writing: Main principles ▾

About Olena Rossi ▾

Newsletter

Contact

Privacy Policy

Terms of use

Item writing terminology

Test Formats

Task types >

Edit

Sign up for updates

Multiple Choice

Multiple Matching

Short-Answer Questions



Learn about most common selected-response

- Avoid using ridiculous or humorous distractors.

My friend has a ____ job at a chemist's.

- A half-time
- B spare-time
- C empty-time
- D part-time

Option C 'empty-time' is implausible as a job type so will be easily rejected even by weak test-takers.

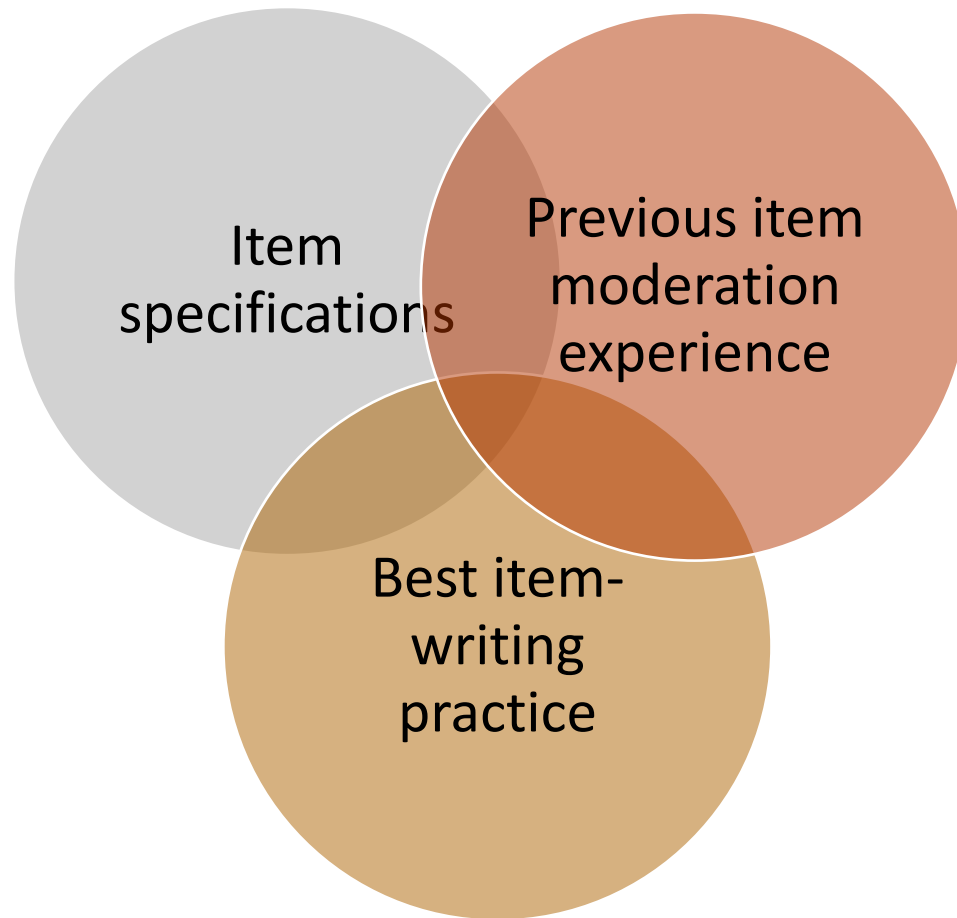


- Avoid using 'All of the above' or 'None of the above' as options.

[Download this page as a worksheet](#)

[Download MC moderation checklist](#)

Item moderation checklists



The final item moderation checklist should draw on all three sources

Giving feedback on items

Useful feedback is

- actionable

- impersonal

- inoffensive

- positive

This is a wrong structure.



The item is supposed to target relative clauses, but the options focus on tenses. Instead of varying the tenses, try varying the relative pronouns and/or word order.



Giving feedback on items

Useful feedback is

- actionable
- impersonal
- inoffensive
- positive

You used a wrong structure.



The item is supposed to target relative clauses, but the options focus on tenses.



Giving feedback on items

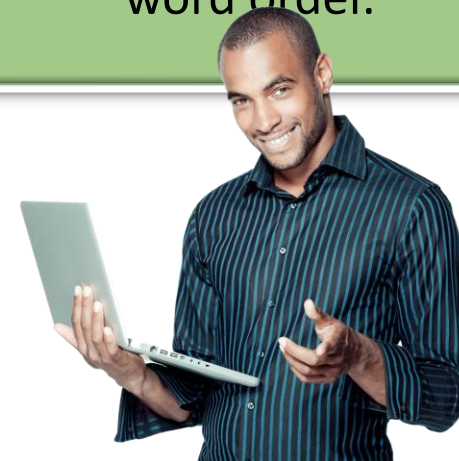
Useful feedback is

- actionable
- impersonal
- inoffensive
- positive

You used a **WRONG** structure again! Please check with the specifications!



The item is supposed to target relative clauses, but the options focus on tenses. Instead of varying the tenses, try varying the relative pronoun and/or word order.



Giving feedback on items

Useful feedback is

- actionable
- impersonal
- inoffensive
- **positive**

This is a wrong structure

The stem is over the word
limit



The stem provides good
contextualisation for the
construct targeted with the
item. And **you** always choose
interesting topics 😊



References

Fulcher, G. (2010). *Practical language testing*. Hodder Education.

Green, A., & Jay, D. (2005). Quality assurance and quality control: Reviewing and pretesting examination material at Cambridge ESOL. *Research Notes*, 21, 5–7.

Hughes A., & Hughes, J. (2020). *Testing for language teachers* (3rd ed.). Cambridge University Press.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10th ed.). Pearson Education Ltd.

**THANK
YOU!**



Olena Rossi

o.rossi@lancaster.ac.uk
olena.rossi@itemwriting.co

