

One case. Four decisions.

A human-in-the-loop validity lab for writing assessment · facilitated by Sahbi Hidri

Everything online	testingexams.com	Room code	TRIAL26
Your rating goes in at	testingexams.com/rubric/	Group speaker	testingexams.com/report/

How the forty-five minutes runs

0–5 min	Opening verdict. Before you see any score, decide what an AI score should be allowed to do: Permit, Review, Delay or Prohibit.
5–15 min	Independent score. Read the response, gather evidence, score all three rubric dimensions and send your rating with your name.
15–31 min	Score on trial. The facilitator reveals the human and machine scores. Your group builds a case: one claim, one quotation, one limitation.
31–45 min	Review and revote. Write a review rule an examiner could actually apply, give your closing verdict, and see how the room moved.

Before you write anything

When you send a rating you are asked for a **name**. That name, your scores and your comments appear on the facilitator's screen, may be shown to the room, and are kept for analysis after the session. Use a label such as "Examiner 4" if you would rather not use your own name. The student response has been anonymised: no name, candidate number, examiner mark or scanned document is included anywhere in this file.

The response

Writing prompt. Some people believe Tunisian educational curricula should focus more on teaching national history than international history. To what extent do you agree?

Anonymised student response, transcribed from handwriting. Read it once before scoring anything.

History is a very important subject in Tunisian educational system. It helps students to know the very important events in the past to understand their effects on what is happening in the present or what can be happen in the future. Some people believe that Tunisian educational curricula should focus more on teaching national history over international history, while in my opinion I think it's very important to focus also on teaching international history like the national history.

Tunisian educational system must focus more on teaching national history over international history as some people think. In fact, national history helps students to discover their roots and how the old Tunisian people were living to feel that past generations made big achievements to make the new ones live peacefully and in a better conditions. It create to the feelings of pride and honors of their nation. Tunisian students can be inspired from the historical national achievements to make similar ones in their social, political, economical lives. Tunisian education can have more information about their country to share them with their foreign friends. They can also understand the causes of the current nation's situation through focusing more on teaching national history over international history by the educational system.

On other side, I think that international history is important as the national one in the actual Tunisian educational curricula. Tunisian students will be able to know a variety of international historical stories and events like US, GB, Russian histories. The students become more intellectual in their knowledge and being opened to other nations. Studying international history helps the students backgrounds while visiting other nations and discovering their monuments and historical places, facts and traditions like when I visited Istanbul and I had the ability to know their ancient political system and monument like the mosque "Suleymaniye". Tunisian student can have the ability to analyse the political, social and economical global conflicts by relating the past events to current ones like the war between Gaza and Israel, so students can link the historical events to current ones.

History is rich of events and facts and people have to navigate into its details. International or national history the important thing is to be intellectual and can be able to understand in the nation's country and international ones in the same level of knowledge.

Spelling, punctuation and grammar are exactly as written. Errors are evidence, not typing mistakes.

SECTION 2

The rubric — 20 points

Dimension	High	Developing	Limited
Content & argument 0–8	7–8 Clear position; relevant reasons are developed.	4–6 Mostly clear and relevant; support is uneven.	1–3 Limited or off-task support. 0 nothing to assess.
Organisation & coherence 0–6	5–6 Clear progression; ideas connect coherently.	3–4 Recognisable structure; connections are uneven.	1–2 Weak sequencing or fragmentation. 0 no evidence.
Language & expression 0–6	5–6 Clear control and sufficient range.	3–4 Meaning generally clear; control is uneven.	1–2 Meaning often obscured. 0 nothing to assess.

Apply each dimension independently. Do not convert the scale, and do not invent criteria.

SECTION 3

Evidence record

Copy the writer's exact words. Two quotations, no paraphrase. If you cannot find words that earn the score, the score is a feeling rather than a judgement.

Quotation 1

Quotation 2

SECTION 4

Your rating

Write it here, then enter the same figures at testingexams.com/rubric/ with your name and the room code. Your online entry is the one that reaches the facilitator's screen.

Dimension	Circle your score	Why — in one line
Content & argument /8	0 1 2 3 4 5 6 7 8	
Organisation & coherence /6	0 1 2 3 4 5 6	
Language & expression /6	0 1 2 3 4 5 6	
Total _____ / 20		Your confidence 1 2 3 4 5 1 very uncertain, 5 very confident

SECTION 5

Micro-courtroom sheet

After the reveal. Build a case your group could defend to a board of examiners.

One claim

What do you say about this score? Something a colleague could disagree with.

One quotation

The writer's exact words that carry your claim.

One limitation

Something your own case cannot account for. A case without a limitation is advocacy, not assessment.

Your group's verdict on the AI score

■ Permit Use the score as it stands	■ Review A human checks it first	■ Delay Gather more evidence	■ Prohibit Not for this decision
---	--	--	--

One speaker per group enters the group's agreed scores and comments at testingexams.com/report/.

SECTION 6

Your exit rule

"Use human judgement" is not a rule. A rule names a trigger, a threshold, an action and a person.

Trigger What condition sets it off?	Threshold How far apart, exactly?	Action What happens next?	Person Who does it?
---	---	-------------------------------------	-------------------------------

A rule that works. "Where the model and the examiner differ by two bands or more on Content, a second examiner reads the script before the result is released."

Closing verdict — after everything you have heard

■ Permit	■ Review	■ Delay	■ Prohibit
-----------------	-----------------	----------------	-------------------

What changed your mind, if anything?
