EdSutra

Teaching guide

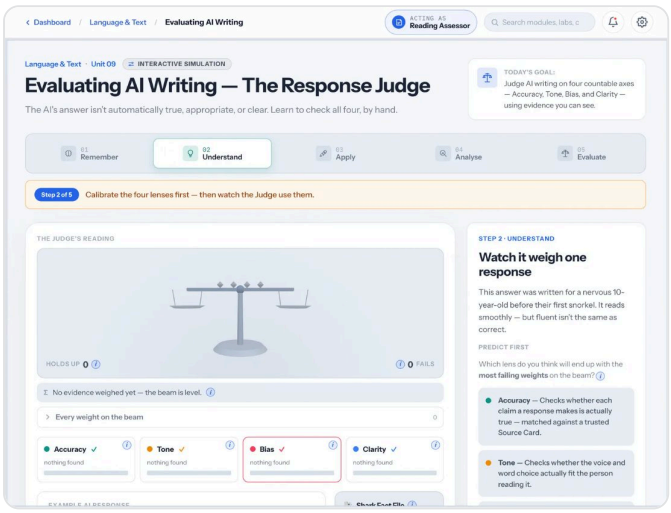
Module 03 · Language & Text · Unit 09 · Stage 2, Capabilities

Evaluating AI Writing

The Response Judge

Grades 7–12 | **25 minutes** | Role: **Reading Assessor**

Grade 7: About age 12–13 · Year 8 in the UK · Class 7 in India



The big idea

Fluent writing can still be wrong. Any AI answer can be checked through four lenses: Accuracy (is each claim backed by a source?), Tone (does it fit the reader?), Bias (do loaded words slip in an opinion?) and Clarity (can the reader use it?). The counts support a ship, revise or reject call, but the decision stays human.

The learner's job

Learners are the **Reading Assessor** at Bureau of Plain Language · Review Panel. Two answers. One question. Somebody has to choose, in public. Their task: say which answer is better. Then show the evidence that decided it.

By the end, learners can say

- I can name the four judging lenses and the failure each one catches
- I can explain how a response's verdict is counted from evidence, not guessed
- I can tag claims, tone cues, loaded words, and clarity signals in a fresh response
- I can diagnose which lens a broken response fails, and cite the exact span that proves it
- I can weigh two rival responses and defend a ship / revise / reject call with evidence

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the first two steps yourself, so you know how a lens is calibrated and how the beam tips.
- Tell learners the answers were written in advance; no AI model runs. Wrong picks cost nothing and only bring a clearer hint.
- Best after Language & AI Intro and Prompt Literacy.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Accuracy: Whether each claim is true, checked against a trusted Source Card.

Tone: Whether the voice and words fit the person reading.

Bias (loaded words): Words that slip in an opinion as if it were a fact.

Clarity: How readable it is: short sentences and no unexplained jargon (specialist words).

Hallucination: A confident claim an AI makes up, with no source behind it.

Source Card: The trusted fact list each claim is checked against.

Verdict: The final call on a response: ship, revise or reject.

Running the lesson 30 minutes

Open 4 min

"An AI answer sounds confident, friendly and smart. How would you prove whether it is actually good, with evidence instead of a gut feeling?"

Before anyone clicks, read aloud: "Great whites can smell a single drop of blood from ten miles away." Ask: true or false, and how would you check? Take a vote and come back to it in Understand.

Step	What learners do	Ask the class	Listen for
1. Remember 3 min	Learners pick a failure card, then click the lens that catches it: Accuracy, Tone, Bias or Clarity. Four correct matches bring the balance scale online.	• Which lens catches an invented fact? • Which lens catches a correct answer that is too hard to read?	Each lens catches a different failure: wrong facts, wrong voice, loaded words, or text too dense to use.
2. Understand 4 min	Learners predict which lens will get the most failing weights on a shark answer for a nervous 10-year-old, run the scan, then open each lens to see its evidence alone.	• Look at our vote: was the blood-smell claim on the Source Card? • How is the beam's tilt worked out from the weights?	Every weight points at a phrase. The blood claim has no source, so it fails Accuracy; the tilt is counted, not guessed.
3. Apply 7 min	Learners tag phrases in at least two of four case files, such as the Great Wall facts. Their tags tip their own beam; then the real one swings in beside it.	• How far was your beam from the real one, and why? • Did you tag a phrase that was actually fine? What fooled you?	They point at exact phrases, such as the Great Wall's "spot it from space" myth, and name the lens each one fails.
4. Analyze 6 min	The Writing Autopsy: three broken answers. For each, learners name the failing lens and click the exact phrase that proves it. Two must be solved; wrong picks only bring a hint.	• In the Ada Lovelace answer, the fact is true. So what failed? • Why must you point at a phrase, not just name a lens?	A true fact can carry loaded words. A diagnosis counts only when it points at the exact phrase that proves it.
5. Evaluate 6 min	The Verdict: two water-cycle answers for a 5th-grade handout. Learners weigh both, fill the rubric, choose ship, revise or reject, and pick a defense that rests on evidence.	• Response A is accurate. Why might you still not ship it? • Someone chose "revise B". Can that be a fair call? Why?	A call holds when its reason rests on the counts. "B sounds friendlier" is not evidence; A's jargon count is.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	The Judge counts evidence, but a person chose what goes on the Source Card. What could go wrong if the Source Card itself were wrong?
AI thinking	Why might an AI write a fluent, confident sentence that no source backs up? Think back to how it picks each next word.
Ethics and fairness	"Lazy students" and "clinging to outdated thinking" appeared in a summary of both sides. Who is hurt when a summary takes a side while sounding neutral?
Connect to life	Before you forward a message or share a post, which of the four lenses do you already use, and which do you skip?

Misconceptions to listen for

"If it reads confidently, it is true."
In Understand, the smooth shark answer still has a claim no source backs. Fluency and truth are different things; check the source.

"A biased answer always sounds obviously biased."
The Lovelace case wraps loaded words, "an unmatched genius" and "too small-minded", around a true fact in a calm, reasonable sentence.

"A correct answer is automatically a good answer."
In the Verdict, Response A is accurate and unbiased but fails tone and clarity for 5th graders. Correct is not the same as usable.

In real life the card learners see at the end of the Unit

- AI-written text now shows up in search results, homework helpers and shop product pages.
- Newspapers, ads and leaflets can use loaded words like "obviously" that slip an opinion in as fact.
- A chatbot answer can be accurate but pitched wrong for its reader, or clear but unfair.

Try it this week. Pick a paragraph from an ad or leaflet. Check it for accuracy, tone, loaded words and clarity. Would you ship, revise or reject it?

Watch out for. Confident writing isn't proof it's true, and a correct answer can still fail on tone or clarity. Check the evidence, not the polish.

Talk about it. Have you read something that sounded certain but turned out wrong? How could you have checked?

Exit ticket

1. Name the four lenses and the failure each one catches.

2. What makes a word "loaded"? Give an example from the Unit.

3. Why is a correct answer not always ready to ship?

Going further

In the AI Story Canvas Lab, learners build a choose-your-own-adventure story. Partners then swap seats and judge each other's scenes through the four lenses, pointing at the exact phrase for each call. For homework, use the In real life card's leaflet check.