



Carleton
University

AI-Assisted Marking with Humans-in-the-Loop

A transparent, auditable workflow for large-scale engineering assessment

Leonard MacEachern, PhD, P.Eng.(APENS)

Associate Professor & Chair
Department of Electronics, Carleton University



Context and Motivation



1) Teaching context



Fall-term co-instructor for a large engineering course



Approx. 650 students across the course



10 labs per student



2) Resource constraint



Historical TA allocation: 14 to 20 TAs



Actual allocation that term: 10 TAs



I prefer to use TA time for student interaction



3) Scale of marking challenge



Each lab: 5 multiple-choice questions, marked conventionally



Each lab: 5 long-form engineering questions requiring rubric-based judgment



Long-form written responses across the term: about 32,500



Prototype question: can AI assist first-pass marking without surrendering human control?



This was a workflow experiment under real teaching constraints, not a procurement recommendation.

Human-in-the-Loop Marking Workflow



5 multiple-choice questions were already marked automatically; the workflow focused on the 5 long-form questions.

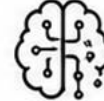
1. Student submission



2. Anonymization by codeword



3. Rubric-based AI first pass



4. Semantic consistency audit



Rubrics were shared after submission

Semantically similar answers with divergent marks were flagged

5. Instructor verification in HTML review interface



6. Selective no-downgrade remarking



7. Transparent release to students



8. Student remark request if needed



Marks were adjusted upward only during consistency review



Architecture is model-agnostic and could be implemented with local models.



Boundaries and Safeguards

Where AI assistance fits

- Best for structured, verifiable responses
- Examples: formulas, derivations, numerical results, fill-in responses, and rubric-based engineering explanations
- Even complicated reasoning can often be checked against an explicit framework
- Reasoning is not the same as thinking
- Poor fit for creative writing or work intended to evoke human emotion

Safeguards for fairness and trust

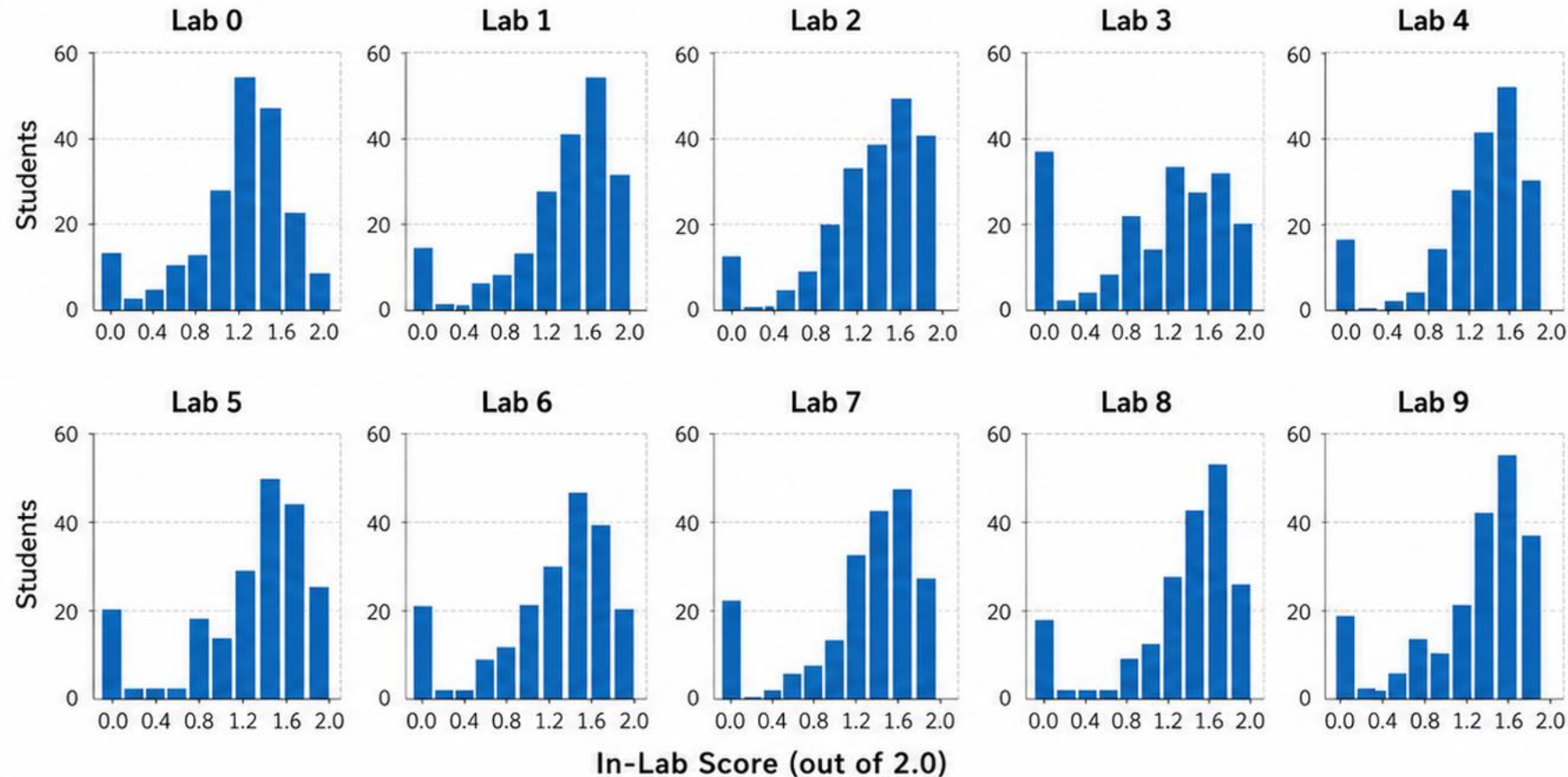
- Anonymized student work
- Detailed rubrics, like those given to TAs
- Independent semantic consistency checks
- Instructor review before release
- No score reductions during consistency re-review
- Transparent student-facing release using codewords
- Remark path for comparable cases



The model mattered less than the surrounding process: anonymization, rubrics, auditing, verification, and transparency.

Section-Level Results: In-Lab Histograms

Pre-lab marks excluded, example section $n = 202$



 Each histogram shows the number of students in each 0.2-wide score bin. Scores are out of 2.0.

Observations

-  Most students cluster between 1.2 and 1.8 / 2.0
-  Lab 3 shows the widest spread and the most zero-score outcomes
-  Later labs shift upward, with Lab 9 showing stronger high-end performance
-  Distribution views help identify difficulty spikes, anomalies, and non-participation

What Students Received

results.engineering.carleton.ca/report

Lab 03 · Section B · Question 1 Final Results

Question: 1 Section: B Generated: 2025-12-01 07:49

Question prompt
When you added resistor R3 (going from Figure 6 to Figure 7) did the voltage across R2 change? Explain why.

Responses 202	Mean score 4.64 / 5	Median 5 / 5	Updated via remark 3
------------------	------------------------	-----------------	-------------------------

Codeword	Answer	Feedback / reasoning	Score
1a1a1a1a	When adding R3 to the circuit, the voltage across R2 changes from 4.420V to 4.100V. This is due to the current divider equation and the nature of ohm's law ($V = IR$). In order to find the current going over	Feedback Clear: reports VR2 change (4.420 V → 4.100 V) and correctly attributes it to current division and Ohm's law causing a reduced voltage across R2 when R3 is added in parallel. Reasoning Student gives numerical before/after values and correctly explains that adding R3 creates a parallel branch so currents	5 / 5
1Atest	Yes the voltage across R2 decreased by about 1/3 of a volt. This change occurs because when you add R3 the equivalent resistance between nodes decreases. This means that the equivalent series circuit would have a	Feedback States VR2 decreased by ~0.33 V and correctly links the drop to lower equivalent resistance, higher series current, larger VR1, and thus less voltage for the lower branch. Reasoning Provides a measured/quantified change and explains via equivalent resistance decrease and voltage-divider effect (higher	5 / 5
5795020170	Yes, it did. In the series circuit, voltage at R ₂ was measured to be 4.26V. In the parallel circuit, R ₂ was measured to be 4.11V. That is because when R3 was added, R2 became parallel with R3. This reduces their	Feedback Provides before/after measurements (4.26 V → 4.11 V) and correctly explains that adding R3 makes R2 R3 lower the branch resistance, shifting the divider and reducing VR2. Reasoning Measured voltages and correct technical explanation: R3 in parallel with R2 reduces the branch equivalent (1800 Ω → 1080 Ω in	5 / 5

Remark request (optional)
If you believe there is an issue with your evaluation, you can submit a quick remark request.

[Request a remark](#)

Anonymized by
student-chosen
codeword

Exact **student
answer** shown

Students could **compare
anonymized cases** for
consistency

Exact **question
prompt** shown

Updated-via-remark
tags where
applicable

Rubric-referenced
feedback and
reasoning

Final **score** clearly
displayed

A simple form allowed
remark requests

Transparency was used as a fairness mechanism, not just a reporting feature.



What I Learned

What worked

- AI assistance can work for bounded, rubric-based engineering marking
- The gain comes from shifting work from first-pass marking to faster human verification
- A cautious semantic audit can surface potential inconsistencies before release
- Transparency and clear rubrics improved trust and auditability

Caveats

- Not suitable for all forms of assessment
- Human judgment remains essential
- The process must include appeal or remark pathways
- Local institutional models may be preferable for governance reasons
- The surrounding workflow matters at least as much as the model



My conclusion: AI-assisted marking is viable for structured assessments when humans remain in control.

