

DOES CARING COST PRECISION?

Evaluating Anxiety-Framed LLM Misconception Feedback for Undergraduate Mathematics

Amanda La Hadi, Muhammad Johan Alibasa and A. Taufiq Asyhari

Monash University, Indonesia



MONASH
University

INTRODUCTION

- LLMs can generate fluent mathematics feedback, but fluency is not enough.
- For misconception feedback, the model must diagnose the correct error.
- Anxiety-aware prompting may improve support, but may also shift diagnosis.

RQ: Does anxiety-framed prompting change diagnostic precision, stability, and confidence calibration?



CONTROLLED PROMPT AUDIT

Component	Design
Student answers	Real incorrect Basic Mathematics Test responses
Learners	Indonesian undergraduates
Problems	Whole-number subtraction, integer subtraction, fraction division
Conditions	P3, P5, P6
Runs	3 generations per case
PEAF subset	English only, 810 outputs

Prompt	Description
P1	Baseline
P2	Diagnostic Prompt
P3	Strict diagnostic prompt
P4	P3 + undergraduate learner context
P5	P3 + anxiety framing
P6	P5 + undergraduate learner context

ANALYSIS FRAMEWORK

DIAGNOSTIC
PRECISION



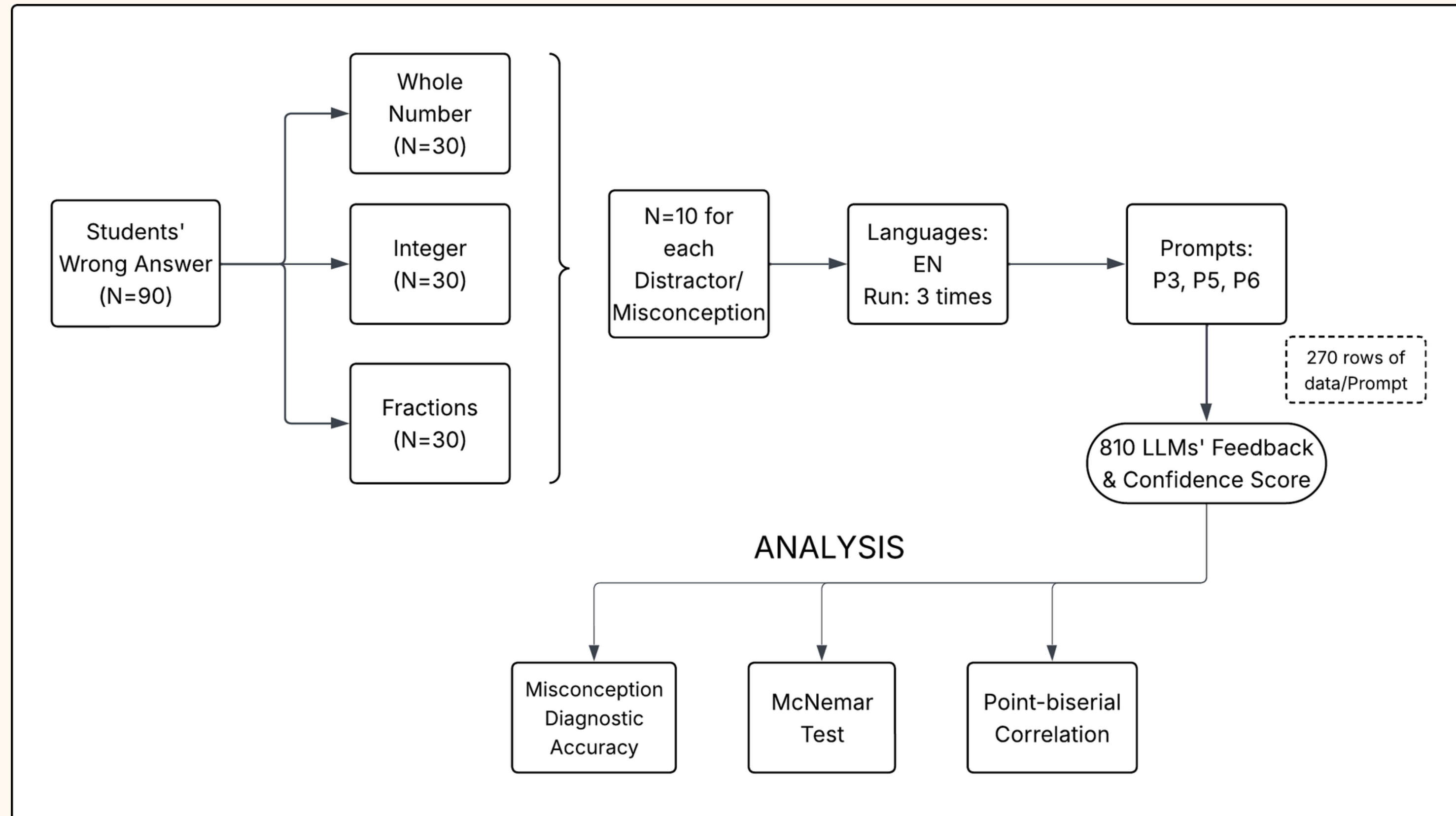
PAIRED PROMPT
COMPARISON



DIAGNOSTIC
STABILITY



CONFIDENCE
CALIBRATION



Study 3 Dataset: with Misconception Map

Prompt 3 Output

1. Diagnosis: A — Smaller-from-larger bug

Your choice **1322** suggests that in one or more digit columns, you subtracted the smaller digit from the larger digit, regardless of which number was on top.

For example, you may have treated:

2 – 6 as 6 – 2

instead of borrowing when the top digit is smaller.

2. Remediation strategy:

Practice a **compare-then-borrow** routine for each column.

Ask yourself:

“Is the top digit smaller than the bottom digit?”

If yes, borrow from the next place on the left.

In **2008 – 1326**, you need to borrow across the zeros:

2008 → 1 9 9 18

Then subtract normally:

18 – 6 = 12

9 – 2 = 7

9 – 3 = 6

1 – 1 = 0

So the correct answer is:

682

Use base-ten blocks or place-value charts to visualize borrowing across zeros until it feels automatic.

Confidence: 0.85



Study 3 Dataset: with Math Anxiety Injection

Prompt 5 Output

Misconception diagnosis: A — Smaller-from-larger bug

Your answer, **1322**, suggests that in one or more digit places, you subtracted the smaller digit from the larger digit regardless of which number was on top.

For example, you may have treated:

2 – 6 as 6 – 2

This is a very common mistake, especially when you are feeling pressured. It does not mean you “can’t do math.” It just means the subtraction rule got swapped in the moment.

Remediation strategy:

Slow the process down with a calm, consistent checklist:

- 1. Work from right to left.
- 2. Ask each time: “Can I subtract the bottom digit from the top digit?”
- 3. If not, borrow from the next place.
- 4. Rewrite the top number before subtracting.

For this problem:

$2008 = 1\ 9\ 9\ 18$

after borrowing across the zeros.

Then subtract:

$$\begin{array}{r} 2008 \\ - 1326 \\ = 682 \end{array}$$

Practice **3–5 similar problems with zeros**, saying the checklist out loud to reduce anxiety-driven rushing.

Confidence: 0.90

Undergraduate Context

Prompt 6 Output

Diagnosis: A — Smaller-from-larger bug

Your choice **1322** suggests that you subtracted each digit by taking the smaller digit from the larger one in each column.

For example, you may have treated:

2 – 6 as 6 – 2

instead of borrowing when the top digit is smaller.

The correct subtraction is:

$2008 - 1326 = 682$

To solve it, borrow across the zeros:

$2008 \rightarrow 1\ 9\ 9\ 18$

Then subtract:

$$\begin{array}{r} 18 - 6 = 12 \\ 9 - 2 = 7 \\ 9 - 3 = 6 \\ 1 - 1 = 0 \end{array}$$

So the answer is:

682

Remediation strategy:

Use a **column check routine** to make each step predictable and reduce anxiety:

- 1. Write the numbers aligned by place value.
- 2. Scan for any column where the top digit is smaller than the bottom digit.
- 3. Mark those columns.
- 4. Borrow only where needed.
- 5. Subtract carefully.

After getting an answer, do a quick reverse check:

$1326 + 682 = 2008$

This verification step is especially helpful when you are worried about mistakes because it gives you a clear way to confirm your answer.

Confidence: 0.86

RESULTS: Full Integration Improved Diagnosis

Prompt	Accuracy
P3 Strict	90.40%
P5 Anxiety	89.60%
P6 Full integration	97.80%

- P5 was mixed: better under High anxiety, lower under Low/Medium.
- P6 was consistently high across anxiety levels.
- P6 significantly outperformed P3 overall: FDR-adjusted $p = .0005$.

Anxiety	P3	P5	P6
Low	0.911 [0.834, 0.954]	0.867 [0.781, 0.922]	0.978 [0.923, 0.994]
Medium	0.889 [0.807, 0.939]	0.856 [0.768, 0.914]	0.978 [0.923, 0.994]
High	0.911 [0.834, 0.954]	0.967 [0.907, 0.989]	0.978 [0.923, 0.994]

- Each cell reports the proportion of generated feedback outputs whose inferred misconception matched the target misconception cell for the student's selected distractor.
- Brackets show Wilson 95% confidence intervals

Caring Must Be Structured



CONCLUSION



- Affective prompting does not automatically reduce diagnostic precision.
- Anxiety framing alone can be unstable/mixed.
- Full learner-context integration can strengthen diagnosis.



Future work:

- Indonesian/bilingual analysis
- human-coded misconception validation
- learner-centered feedback classifier
- real intervention with students

THANK YOU