

EDULEARN 2026

Generative AI for Multiple Choice STEM Assessments

CHRISTINA SPIROU PERDIKOULIAS

CHAD VANCE

DR. STEPHEN WATT

© DigitalEd, 2018-2026. Möbius is a trademark of DigitalEd.

UNIVERSITY OF
WATERLOO



DigitalEd

WHY IS MCQ AUTHORIZING DIFFICULT?

*High-quality multiple-choice questions depend heavily on **well-crafted distractors**, which remains a difficult and time-intensive task*

- Creating plausible distractors is the most difficult part of MCQ design
- Higher-order STEM questions are difficult and time-consuming to author
- Effective distractors reflect authentic misconceptions, rather than random errors
- Poor distractors reduce both question quality and assessment validity

*The goal isn't to generate a wrong answer;
It is to generate plausible, misconception-driven incorrect answers
that reveal how students think*

THE AI OPPORTUNITY

*Generative AI struggles with **mathematical reasoning***

Research shows that LLMs can:

- make procedural errors
- apply concepts and rules incorrectly
- follow flawed reasoning paths
- generate answers that appear plausible but are mathematically incorrect

*These same error patterns often **mirror authentic student misconceptions**, making them valuable for distractor generation*

WHY MULTIPLE CHOICE WORKS

MC assessments provide a *structured format* that makes AI-generated content *easier to validate and deploy*

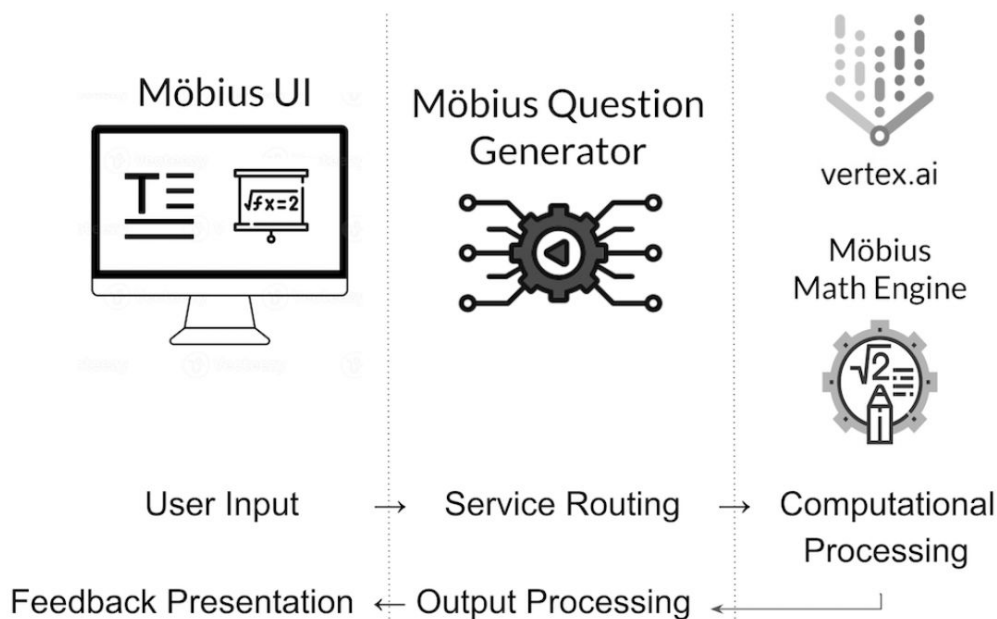
Multiple choice provides the following:

- constrained output space
- finite answer set
- easier validation
- automated checking

In STEM, selecting the correct answer typically requires completing the *underlying mathematical or scientific reasoning process*

SYSTEM ARCHITECTURE

Combining generative AI with a computation engine enables *scalable question generation* without sacrificing mathematical rigor



PRESERVING MATHEMATICAL MEANING

*Maintaining mathematical semantics ensures AI-generated content remains **interpretable, evaluable, and instructionally accurate***

CHALLENGE

Mathematics is not plain text

SOLUTION

Use structured mathematical representations:

- MathML
- LaTeX
- semantic mathematical encoding

*Preserving mathematical semantics ensures that AI-generated content remains **meaningful, computable, and verifiable***

VALIDATION LAYER

A *deterministic validation* process transforms AI-generated possibilities into *reliable assessment content*

Generate



question

correct answer

distractors



Validate



unique correct answer

distractors are incorrect

symbolic equivalence

evaluability

```
FUNCTION validateSingleAnswer(question)
// 1. Create an empty list to store answers that are verified as correct
SET correctAnswersList TO new empty List
// 2. Loop through every answer provided in the question
FOR EACH answer IN question.getAnswers()
// 3. Use an external engine to check if this answer is a valid solution for the equation.
IF mobiusMathEngine.validateAnswer(question, answer) IS TRUE THEN
// 4. If it is valid, add it to our list
ADD answer TO correctAnswersList
END IF
END FOR
// 5. After checking all answers, inspect the list of valid solutions
// 6. If the list is empty, no correct answer was found. This is an error.
IF correctAnswersList is empty THEN
THROW Exception("No correct answer found for question: ", question, correctAnswersList)
// 7. If the list has more than one, the question is ambiguous. This is an error.
ELSE IF size of correctAnswersList IS GREATER THAN 1 THEN
THROW Exception("Multiple correct answers found for question: ", question, correctAnswersList)
END IF
// 8. If no errors were thrown, the list must have exactly one answer.
// Return TRUE.
RETURN (size of correctAnswersList IS EQUAL TO 1)
END FUNCTION
```

SIMPLE EXAMPLE

AI can generate **distractors that mirror authentic student misconceptions** while remaining mathematically distinct from the correct answer

Solve:

$$(x + 2)^2 > x^2 + 12x - 20$$

Correct Answer: $x < 3$

Potential Distractors: $x > 3$

$$x < 2$$

$$x > 2$$

Each distractor represents a **realistic mistake** in algebraic manipulation, rule application, or a combination of both

FINDINGS

*When paired with **computational validation**, generative AI can significantly **reduce assessment authoring effort** while preserving **quality***

Generative AI can:

- accelerate assessment authoring
- generate plausible distractors
- support content variation
- reduce instructor effort

When combined with validation:

- maintains rigour
- maintains correctness
- supports repeatability

TAKEAWAYS & FUTURE WORK

*The value of generative AI in STEM assessment may not be in solving problems correctly, but in **helping educators create better ways to test understanding***

TAKEAWAYS

- AI weaknesses can become instructional assets
- Distractor generation is a promising AI use case
- Validation is essential
- Generate + Validate is a practical pattern for STEM assessment

FUTURE WORK

- automated validation suites
- larger-scale evaluation
- commercial deployment
- broader STEM domains

TAKEAWAYS & FUTURE WORK

*The value of generative AI in STEM assessment may not be in solving problems correctly, but in **helping educators create better ways to test understanding***

TAKEAWAYS

- AI weaknesses can become instructional assets
- Distractor generation is a promising AI use case
- Validation is essential
- Generate + Validate is a practical pattern for STEM assessment

FUTURE WORK

- automated validation suites
- larger-scale evaluation
- commercial deployment
- broader STEM domains

thank you