Students' Perceptions to a Large Language Model's Generated Feedback and Scores of Argumentation Essays

Winter Allen, Anand Shanker, N. Sanjay Rebello

Purdue University, West Lafayette, IN, 47906

Introduction

- Learning to define problems, design solutions, and develop scientific arguments are key science and engineering practices identified in the Next Generation Science Standards(NGSS)[1]
- Students were taught to develop an argument based on Claim, Evidence and Reasoning from Mcneill and Krajcik. [2]
- Research has shown that feedback is one of the important drivers of learning. [3]
- By combining the benefits of feedback with generative AI there is potential for providing students with constructive feedback that aids in their learning.

Research Question

To what extent can an LLM score and provide feedback in a way that students find useful and accurate?

Methods

Context

- First semester calculus-based physics course at a large U.S. Midwestern land grant university
- Students tasked with writing a scientific argument (Claims, evidence and reasoning) to support their answer to a problem in an online multiple-choice quiz.

Feedback Generation

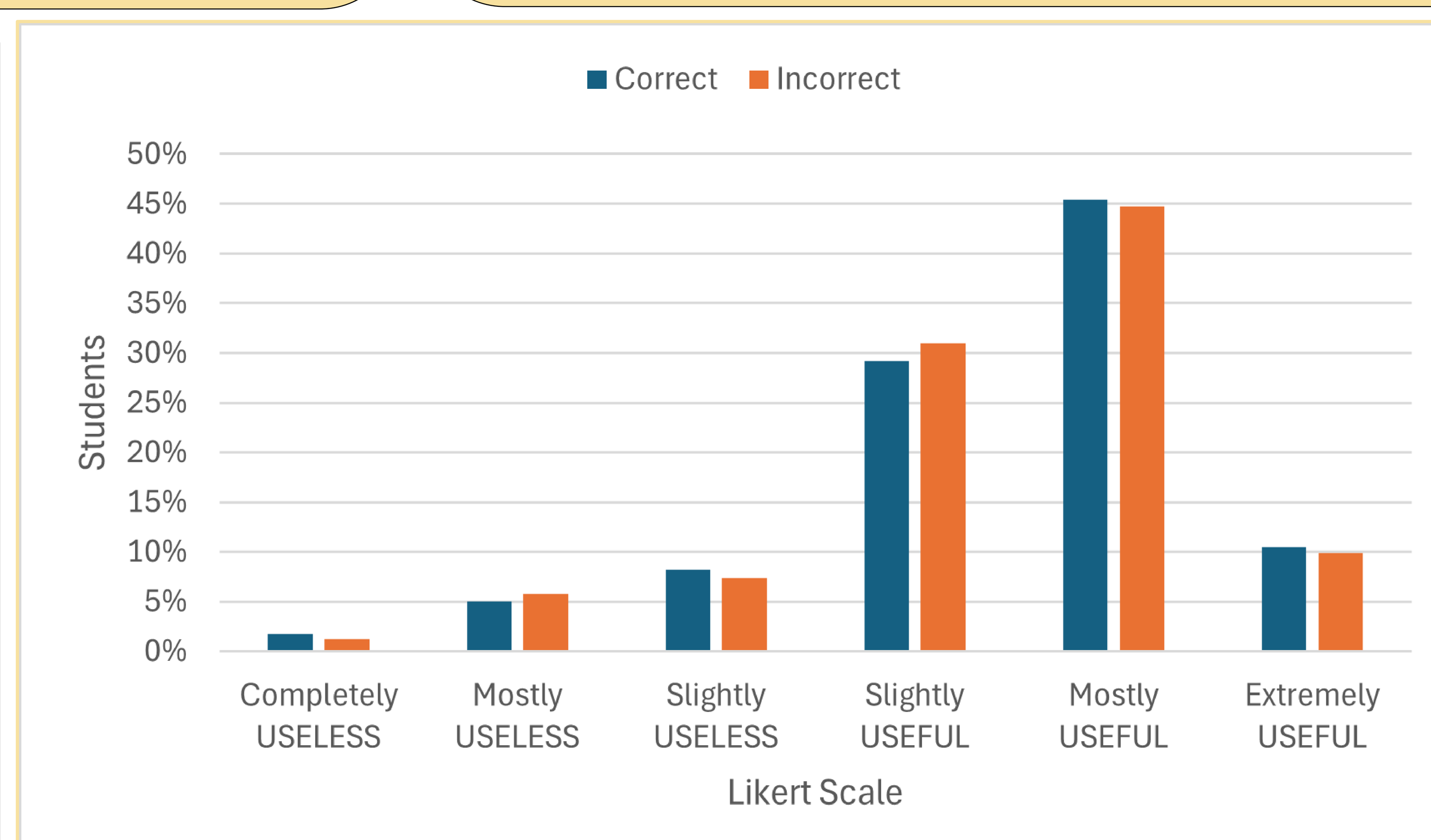
- OpenAI's GPT 4o [4] was prompted to think as an educator evaluating the physics aspects of an essay of any essays over 50 words with the temperature set to 0.80.
- We provided the problem students solved, a 5-point rubric, and an example of an ideal answer, along with discouraging the LLM from focusing on aspects that did not focus on the physics.

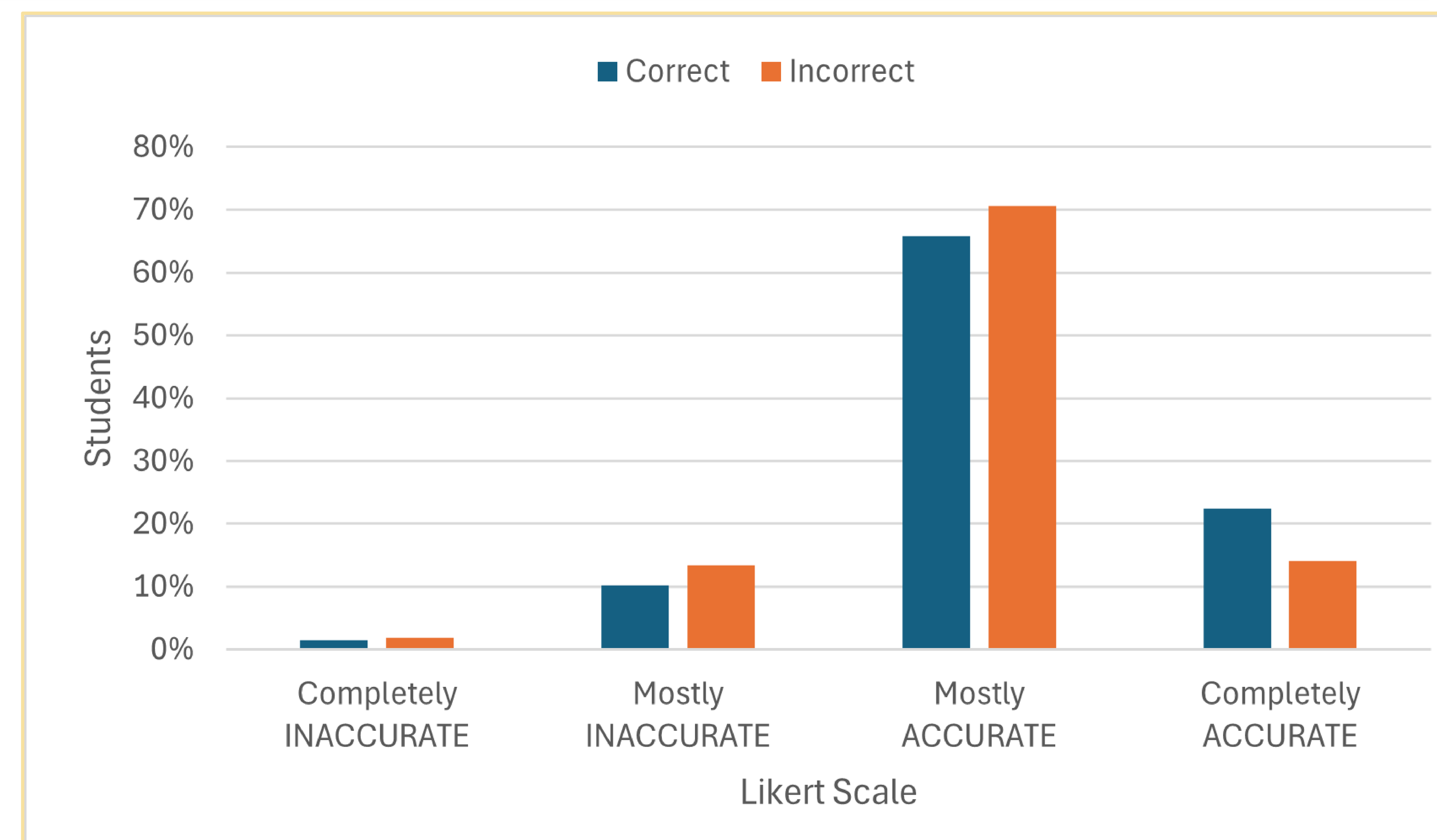
Data Collection

- Students were given the feedback and score via a tiny URL.
- On a post survey, students were tasked with ranking the score and feedback's usefulness on a 6-point Likert scale and the accuracy on a 4-point Likert scale.

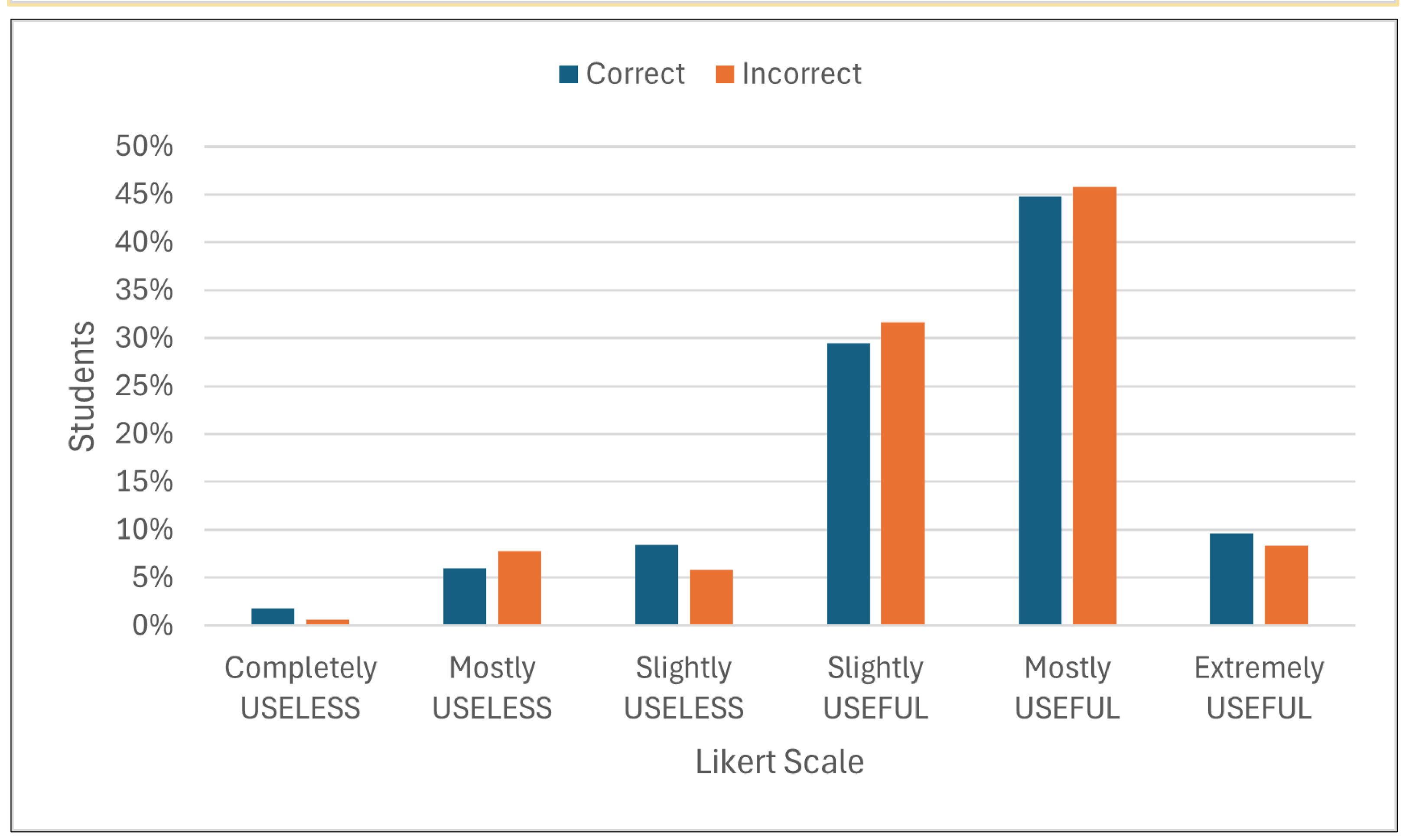
	Quiz 08	Quiz 09
Total Average ± SD	1.69 ±1.29	2.72 ±1.08
Correct Average ± SD	2.14 ±1.24	2.77 ±1.08
Incorrect Average ± SD	1.12 ±1.11	2.53 ±1.04
p-value	<<0.001	0.015
Cohen's D	00.87	0.22

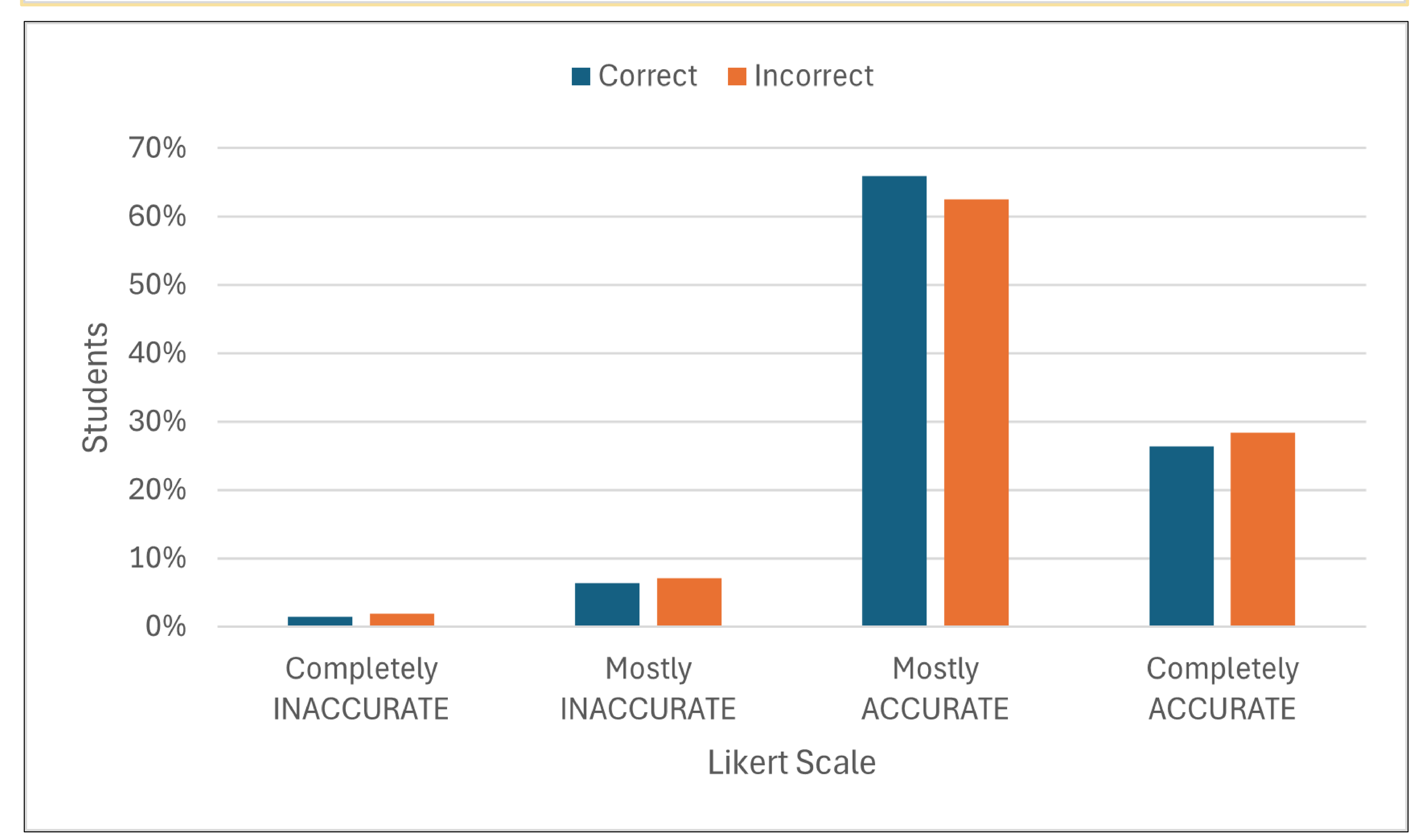
Quiz 08





Quiz 09





REFERENCES

[1] National Research Council, Next Generation Science Standards: For States, By States (The National Academies Press, Washington, DC, 2013).
[2] K. L. McNeill and J. S. Krajcik, Supporting grade 5-8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing., Pearson (2011).
[3] Hattie and H. Timperley, The power of feedback, Review of Educational Research 77, 81 (2007).
[4] OpenAI, Gpt-4o (2024), large multimodal model.

ACKNOWLEDGEMENTS

This research is supported in part by NSF Grants 2300645 and 2111138. Any opinions, results, and findings expressed here are those of the authors and not of the NSF.