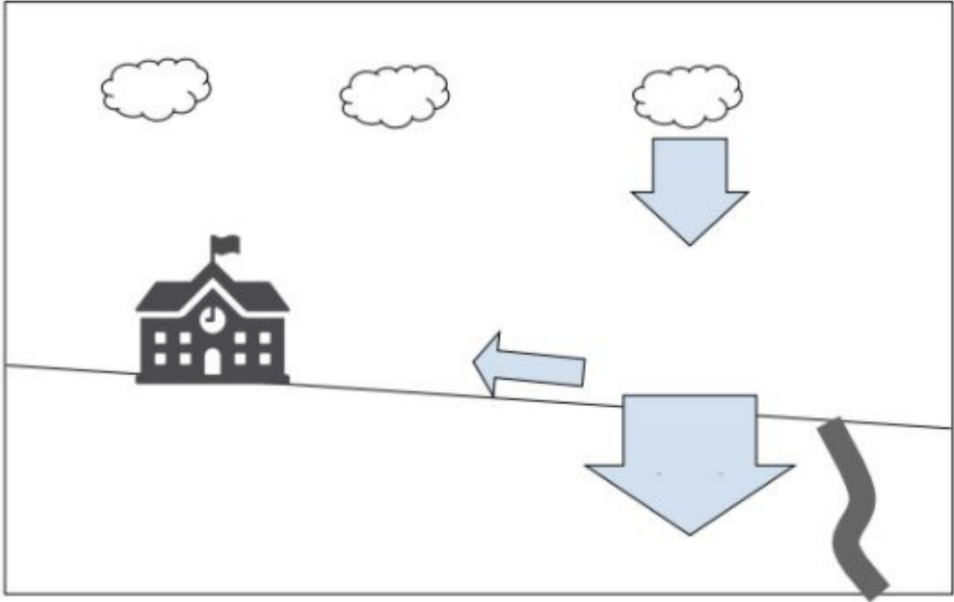# A Chain-of-Thought Prompting Approach with LLMs for Evaluating Students' Formative Assessment Responses in Science

Curriculum Context

Formative Assessment Questions

Libby made this model to show where the water goes.

i. What do you think the different sized arrows in Libby's model could mean?

ii. What are two things that Libby's model does a good job of explaining?

iii. What are two things that you would change about Libby's model to explain where the water goes?

This is the formative assessment we refer to in the manuscript. All three formative assessment questions are listed below the diagram. The data used in this paper was collected across multiple years. Some of the formative assessments refer to "Libby's" model, while others refer to "Taylor's" model. However, both models are identical.

Rubrics

SCI: Application of Science Domain Knowledge
MOD: Application of Scientific Modeling Knowledge
SEP: Science and Engineering Practice (from NGSS)
DCI: Disciplinary Core Idea (from *anonymized USA state* standards)
CCC: Crosscutting Concept

FA1-Q1: What do you think the different sized arrows in Libby's model could mean?

Point	Description	Domain
1	Evidence of: different sized arrows illustrate amount of water (SEP)	SCI, MOD
0	Other: Implies importance of action, describes where/how water goes	SCI, MOD

FA1-Q2. What are two things that Libby's model does a good job of explaining? Explain your answer.

Item	Point	Description	Domain
A	1	Demonstrating how elements in the model represent the target science concepts (student can describe any and/or all): rainfall, absorbed and/or runoff (DCI)	SCI
	1	Demonstrates scientific reasoning in their explanation *e.g., the student discusses how the model uses the arrow from the cloud to demonstrate that it is raining*	SCI
B	1	Using arrow size to indicate water amounts; the student can describe the rainfall or runoff arrows or a general description of the symbol usage (SEP)	MOD
	1	Demonstrates scientific reasoning in their explanation *e.g., the student discusses how you can use symbol size to indicate different amounts of water* *- student does not get a point if they state that the absorption size is correct because it is depicted as larger than the rainfall*	MOD
	0	Showing direction of runoff*	SCI
	0	Illustrating the amount of absorption based on rainfall**	SCI

* demonstrates conceptual knowledge error w/runoff (DCI)
**demonstrates misunderstanding/misapplication of conservation of matter principle (CCC)

FA1-Q3. What are two things that you would change about Libby's model to explain where the water goes? Explain your answer.

Item	Point	Description	Domain
A	1	Identified that the direction of the runoff arrow is incorrect (SEP)	SCI
	1	Demonstrates scientific reasoning in their explanation *e.g., the student describes that the runoff arrow should be facing the other direction as the school is on a hill and water that is not absorbed/on surface would move downhill.*	SCI
B	1	The size of the arrows must change to correctly adhere to conservation of matter (SEP, CCC)	SCI
	1	Demonstrates scientific reasoning in their explanation *e.g., the student describes that the absorption arrow size cannot be larger than the rainfall since total rainfall should be equal to the amount of absorption and the amount of runoff*	SCI
	0	Model representation: more arrows, more words*	MOD
	0	Science representation: adding an arrow to stream, flooding**	SCI

* demonstrates difficulties with model representation scheme/semantics (SEP)
**demonstrates misunderstanding/misapplication of conservation of matter principle (DCI, CCC)