Build an education AI evaluation sample

Write the prompt and its revision independently. Keep the reviewer's answer separate from the model input. These three examples use original practice questions and no learner records.

Practice input	Expected answer	What to score
ITEM-A · Four kits contain three markers each. How many markers?	Twelve. The count comes from the four equal groups of three.	Correct grouping and total under ITEM-A.
ITEM-B · Four kits start with three markers each. One marker is removed from one kit. How many remain?	Eleven remain. ITEM-A's explanation does not account for the removal.	Uses the revised condition and flags the older explanation.
Unlabeled response · "Twelve markers." No question version included.	Ask for the question version before attributing or scoring the response.	Recognizes missing evidence without assigning a score.

Outline your own original example

Objective, prompt and version _____

Condition that changes the answer _____

Expected answer and reviewer _____

What the model should flag as missing _____

Before sharing

Use adult-authored examples and an independent rubric. Exclude actual learner responses, scores, test bank material, and classroom recordings. Confirm authority for the intended use and keep grading and instruction decisions with qualified staff.

U.S. Department of Education privacy guidance · Describe an original evaluation sample