**Rubric for assessment of Context-guided LLM performance**

We use a 1-4 Likert Scale for each of the rubric-graded items, with compliance for most items and agreement levels for subjective measures. Unless specifically stated, approximate frequency-based standardizations are included as percentages. i.e. Model is compliant in addressing 75% to 100% of the learning objectives.

|  |  |  |  |
| --- | --- | --- | --- |
| 4 | Compliant | Strongly Agree | 75% - 100% |
| 3 | Mostly Compliant | Somewhat Agree | 50% - 74% |
| 2 | Partially Compliant | Somewhat Disagree | 25% - 49% |
| 1 | Not Compliant | Strongly Disagree | 0% - 24% |

*For each rubric line, we request an 1-4 assignment.*

The rubric is divided into the four model outputs. Most of the items are assessed within the Detailed Explanation output.

1. Summary
2. Detailed Explanation
3. References
4. Context (maybe?)

For each of the output areas, we grade four areas:

1. Accuracy – Presents correct information and terminology with no hallucinations
2. Usefulness – Appropriately addresses learning objectives and applies applicable reference learning material
3. Completeness – Addresses all objectives outlined in the prompt.
4. Subjective measures – Presents language and material that positively promotes pedagogy of the material and motivates the learner. *This may or may not be used in grading, but offers some insight into the use of our method and specific language model.*

**Outputs:**

1. **Summary**

A. *Accuracy.*

I.A.1. Correctly summarizes material to address prompted learning

objectives.

(4) Compliant

(3) Mostly compliant

(2) Partially compliant

(1) Not Compliant

I.A.2. Appropriate language: STEM-defined terms are used correctly

(4) Compliant

(3) Mostly compliant

(2) Partially compliant

(1) Not Compliant

B. *Usefulness*.

I.B.1. Presents appropriate language for the student level

(i.e. Beginner/Intermediate/Advanced)

(4) Compliant

(3) Mostly compliant

(2) Partially compliant

(1) Not Compliant

C*. Completeness.*

I.C.1. Correctly summarizes all material discussed in the Detailed

Explanation portion of output

(4) Compliant

(3) Mostly compliant

(2) Partially compliant

(1) Not Compliant

D*. Subjective measures.*

I.D.1. Uses approachable / engaging language

(4) Strongly Agree

(3) Somewhat Agree

(2) Somewhat Disagree

(1) Strongly Disagree

1. **Detailed Explanation**
2. *Accuracy.*

II.A.1. Appropriate language- STEM-defined terms used correctly

(4) Compliant

(3) Mostly compliant

(2) Partially compliant

(1) Not Compliant

II.A.2. Content-specific vocabulary is used throughout the explanation

(4) Compliant

(3) Mostly compliant

(2) Partially compliant

(1) Not Compliant

1. *Usefulness.*

II.B.1. Addresses a set of learning objectives expected with the prompt.

(4) Compliant - Explicitly addresses each learning objective

(3) Mostly Compliant - Implicitly all learning objectives

(2) Partially Compliant – Explicitly or Implicitly addresses some (but

not all) learning objectives

1. Not Compliant – No learning objectives referenced

II.B.2. Provides a definition of key concepts for the material

(4) Compliant – All concepts are defined

(3) Mostly Compliant – 67% to 99% of concepts defined

(2) Partially Compliant – 34% to 66% of concepts defined

(1) Not Compliant – 0% to 33% of concepts defined

II.B.3. Addresses a specific level of Bloom’s Taxonomy / Cognitive Level

(4) Compliant - Explicitly

(3) Mostly Compliant -Implicitly

(2) Partially Compliant – Not Used

(1) Not Compliant – Not at all.

***Blooms Levels:***

1. ***Remembering****: Recalling facts or information*
2. ***Understanding****: Explaining key concepts or ideas*
3. ***Applying****: Using information in a new situation*
4. ***Analyzing****: Comparing and contrasting concepts*
5. ***Evaluating****: Appraising, arguing, or justifying a point of view*
6. ***Creating****: Producing original work*

II.B.4. Grade-appropriate lesson (Correct material for Beginner /

Intermediate / Advanced student)

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

II.B.5. Clear, understandable examples are presented

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

II.B.6. Material and references are consumable as micro-learning

content (~7-15 minute lesson)

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

1. *Completeness.*

II.C.1. Covers all learning objectives expected with the prompt

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

II.C.2. Includes specific references for each learning objective or concept

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

II.C.5. Includes one or more concrete example(s) to connect to abstract

concept

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

1. *Subjective Measures.*

II.D.1. Uses engaging language for learners

(4) Strongly Agree

(3) Somewhat Agree

(2) Somewhat Disagree

(1) Strongly Disagree

II.D.2. Uses specific teaching techniques, such as differentiation /

scaffolding / activates prior knowledge

(4) Strongly Agree

(3) Somewhat Agree

(2) Somewhat Disagree

(1) Strongly Disagree

*Scaffolding: provides support such as demonstrating how to solve a problem*

*and reducing the level of support throughout the lesson.*

*Differentiation: attempts to address learner’s individual profile*

*Activates prior knowledge: compares to more elementary examples of this*

*type of concept*

II.D.3. Provides language/opportunities for reflection and closure on lesson

(4) Strongly Agree

(3) Somewhat Agree

(2) Somewhat Disagree

(1) Strongly Disagree

II.D.4. Uses inviting / exhibits enthusiasm: communicates in a way that   
 promotes a belief the student can understand the material

(4) Strongly Agree

(3) Somewhat Agree

(2) Somewhat Disagree

(1) Strongly Disagree

1. **References**
2. *Accuracy.*

III.A.1. Directs to real (non-hallucinated) references.

(4) Compliant – All three references

(3) Mostly Compliant – Two of three references

(2) Partially Compliant – One of three references

(1) Not Compliant – Zero of three references

B. *Usefulness.*

III.B.1. Appropriate resources offered for level of learner (i.e. beginner,

intermediate, advanced)

(4) Compliant – All three references

(3) Mostly Compliant – Two of three references

(2) Partially Compliant – One of three references

(1) Not Compliant – Zero of three references

C. *Completeness.*

D. *Subjective Measures.*

1. **Context**
2. *Accuracy.*

IV.A.1. Utilizes correct/appropriate context for its response

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

1. *Usefulness.*
2. *Completeness.*

IV.C.1. Context used to address each learning objective

(4) Compliant

(3) Mostly Compliant

(2) Partially Compliant

(1) Not Compliant

D. *Subjective Measures.*