

# Interview Guide – Prompt Evaluation

This interview guide helps evaluate the effectiveness of our current prompt design, using the updated difficulty scale (Beginning → Intermediate → Advanced). The goal is to assess how well the prototype meets teacher needs and aligns with the 10 Gold Standards for prompt design.

## 1. Gold Standards for Prompt Quality

1. Generates questions with increasing difficulty (Beginning → Intermediate → Advanced). 2. Adapts to the student's needs, strengths, and difficulties. 3. Ensures the AI asks questions — not answers. 4. Promotes student thinking, effort, and engagement. 5. Produces concise, clean, ready-to-use output. 6. Uses age-appropriate and context-appropriate language. 7. Ensures content that is safe, accurate, and respectful. 8. Provides clearly sectioned, teacher-friendly formatting. 9. Connects questions to student progress. 10. Produces predictable and consistent output formatting.

## 2. Prototype Evaluation

### A. First Reaction

- What was your first impression of the tool?
- Was it easy to understand what to enter?
- Did the workflow feel intuitive?

### B. Difficulty Levels

- Were the levels clearly distinguishable (Beginning → Intermediate → Advanced)?
- Did the difficulty progression feel appropriate?
- Was anything too easy or too hard?

### C. Adaptation to Student Profile

- Did the output reflect the student's strengths or challenges?
- Did it feel personalised?

### D. Engagement & Thinking

- Would these questions make your student think?
- Could they help maintain engagement?

### E. Output Format & Clarity

- Was the output clear, structured, and ready to use?
- Are the sections easy to scan?

### F. Language & Appropriateness

- Was the language suitable for the student's age and level?
- Was anything confusing or too complex?

### **G. Safety & Accuracy**

- Did the content feel accurate, safe, and respectful?
- Did anything feel biased or inappropriate?

### **H. Progress Connection**

- Can you see how the questions reflect learning progression?
- Would this help assess development?

### **I. Predictability & Workflow Fit**

- Was the format predictable and consistent?
- Would this fit your planning workflow?
- What needs improvement?

### **J. Suggestions for Improvement**

- What should be improved in the prompt?
- What additional inputs would you like?
- What would make the tool more valuable?

### **K. Closing**

- Anything else you'd like to add?
- Would you test the refined version later?