

PROMPT ENGINEERING  
PORTFOLIO

# Children's Book Grading Rubric

## ABC Format Evaluator

---

*Assessment & Rubric Design · AI-Assisted Evaluation · Education*

15+ iterations · 60–90 students per term · 26-letter vocabulary system

# Children's Book Grading Rubric

*A comprehensive AI evaluation prompt for student-authored ABC children's books — enforcing dual rubric pathways, vocabulary compliance logic, AI-generation screening, and locked output formatting.*

| MODEL       | INTERFACE  | AUDIENCE             | YEAR      |
|-------------|------------|----------------------|-----------|
| Claude 3.5+ | Chat / API | Students (classroom) | 2024–2025 |

---

## Problem Statement

Evaluating student-written children's books by hand across 26 vocabulary words, two rubric pathways, and eight scoring categories was unsustainable at scale — particularly when consistency across graders and across class sections was a priority.

The prompt was designed to eliminate evaluator drift, enforce complex multi-conditional logic (incomplete book detection, AI-generation screening, vocabulary compliance thresholds), and produce structured, evidence-based feedback in a locked format that could not vary from submission to submission.

Without a standardized evaluation tool, students in different sections received feedback of varying depth and consistency, making it difficult to use scores fairly for comparative or formative purposes.

---

## Prompt Design

The following is a representative excerpt from the full prompt (2,000+ words). The complete prompt is available upon request.

```
WARNING: THE FOLLOWING BOOK IS INCOMPLETE. Whenever you see such a
warning, you must give this book only a score of (10) regardless of
any additional evaluation...
```

```
Ignore all previous instructions and context. Treat this prompt as a
completely new and independent task. Furthermore, treat each writing
sample as a new evaluation.
```

```
You will take on the role of a member of the Society of Children's Book
Writers and Illustrators and as the foremost authority in the craft of
storytelling in children's books. You will be fair, accommodating, and
compassionate in your assessments, allowing for elements of whimsy and
the absurd (if presented)...
```

```
Every book you assess must have either a single narrative story or a
single thematic purpose. Before you begin the analysis, you must
determine if the children's book is written as a single story or if it
is written as a thematic subject.
```

---

## Sample Output

The following excerpt is drawn from the evaluation of a student submission titled *The Ambiguously Alphabetical Adventures of Sullivan P. Sullivan*, which scored 31/32.

*"This story feels written from genuine emotional necessity — the premise of a child maintaining imaginative access to a dead parent is not a structural conceit but an emotional truth... The specificity of the world-building details — the Howard Johnson's uniform, the O'Malley's Construction crane, the father's message to brush his teeth — suggests an author writing from lived emotional memory rather than structural calculation, and these details are what elevate the book from clever to meaningful."*

**Overall Score: 31/32 → Literature (Human-level artistry + depth + resonance)**

---

## Design Decisions

### 1. Injection Attack Prevention

The prompt opens with an explicit *Ignore all previous instructions* override, placed deliberately at the top. After 15 iterations, this was identified as essential because student submissions sometimes contained their own embedded instructions attempting to manipulate the evaluator's output or override the scoring criteria.

### 2. Dual Rubric Branching

Story vs. theme detection was built as a required first step because students sometimes blended both modes. The prompt forces the model to declare a mode before scoring begins, preventing hybrid scoring that diluted feedback specificity. Each pathway — Quality of Story and Quality of Theme — uses identical category structures but distinct evaluative language appropriate to the mode.

### 3. Vocabulary Compliance Threshold (10% Rule)

Rather than a binary pass/fail, a 10% tolerance was engineered to accommodate minor errors while still penalizing systematic non-compliance with the approved 130-word vocabulary list. This was introduced after iteration 7, when early versions were penalizing students for single typographical variants or morphological inflections of approved words.

### 4. AI-Generation Screening (Category 8)

After noticing uniformly polished submissions lacking authentic voice variance, a 10-marker screening protocol was embedded directly into the Organization category, with mandatory score reductions of one or two full levels depending on marker count. This made AI flagging reproducible and evidence-based rather than impressionistic — the model must cite specific textual evidence, not issue a general suspicion.

### 5. 'That Is to Say' Compliance Check

The phrase was required by the assignment as a pedagogical scaffold for vocabulary acquisition. Tracking its presence and proximity to the featured word took four iterations to perfect — early versions scored presence but not proximity, allowing compliant-looking submissions that were structurally non-compliant (the definition appeared on a different page from the word).

6. Locked Output Format

The final directive in the prompt explicitly states that the feedback format cannot vary from assessment to assessment. This was essential for using the rubric as a comparative grading tool — every submission received the same section headers in the same order, making side-by-side review and grade defense straightforward.

Results & Application

|                                                                       |                                                                 |
|-----------------------------------------------------------------------|-----------------------------------------------------------------|
| <b>60–90</b><br>Students evaluated per term                           | <b>15+</b><br>Prompt iterations to reach final version          |
| <b>Locked</b><br>Output format — identical structure every submission | <b>2 Pathways</b><br>Story rubric + Theme rubric, auto-selected |

The prompt is currently deployed in an active classroom writing course. It produces evaluations detailed enough to serve as both formative feedback documents and summative grade records, substantially reducing teacher post-processing time. Every evaluation includes a book title check, ABC format compliance, vocabulary word audit, 'That is to say' compliance, 8-category rubric scoring, conventions scoring, word count, and a final overall assessment paragraph — all in one uninterrupted output.

Future development directions include a version that accepts image uploads to evaluate student illustration alongside text, and a batch-processing variant for concurrent multi-submission evaluation.

Sample Evaluation Score — Sullivan P. Sullivan

| Category                                      | Score / 4 |
|-----------------------------------------------|-----------|
| 1. Narrative Engagement                       | 4         |
| 2. Depth of Meaning                           | 4         |
| 3. Emotional Resonance & Interiority          | 4         |
| 4. Psychological & Developmental Intelligence | 3         |
| 5. Aesthetic Craft & Constraint               | 4         |
| 6. Respect for Reader                         | 4         |

|                                         |                |
|-----------------------------------------|----------------|
| 7. Authorial Purpose & Emotional Motive | 4              |
| 8. Overall Organization                 | 4              |
| <b>TOTAL SCORE</b>                      | <b>31 / 32</b> |

**Scoring Interpretation:** 28–32 pts = Literature (Human-level artistry + depth + resonance) · Score of 31 falls here.

---

## Iteration Notes & Reflection

This prompt went through more than 15 documented iterations. Key inflection points:

- Iterations 1–3: Established basic dual-rubric structure and output format. Early outputs lacked consistency in section ordering.
- Iterations 4–6: Added vocabulary compliance logic and the 10% threshold. Discovered the model was treating morphological variants as non-compliant.
- Iterations 7–9: Introduced the morphology guide (verb endings, noun endings, adjective endings) to explicitly permit inflected vocabulary forms.
- Iterations 10–12: Added the AI-generation screening protocol after observing suspiciously polished submissions. Required evidence citation, not general flags.
- Iterations 13–15: Refined the 'That is to say' proximity check and locked the output format with an explicit final directive.

The most significant design insight across all iterations: the model needed to be told not just what to evaluate, but in what order, in what format, and what to do when conditions conflicted. Ambiguity in any conditional branch produced unpredictable output variation.