

## Appendix E: Building a Rubric

---

Chapter 5 recommends requiring students to submit a GenAI Appendix and grading it. That recommendation immediately raises an obvious follow-up question: grade it on what, exactly?

This appendix is a faculty resource. It walks you through building a rubric for that purpose, explains what to look for at each performance level, and provides a finished rubric you can use or adapt. The process helps understand how the rubric was built so it can help you adjust it for your course and assignment type. The finished product appears after the instructions.

### What a Rubric Does (and Why It Matters Here)

I know you know this, but to be thorough, a rubric is a scoring guide that breaks a complex task into distinct, evaluable criteria and describes what a submission looks like at each performance level. Rubrics help maintain consistency in grading because you apply the same standard to the first appendix you grade and the last, and students know in advance what “Excellent” means in concrete terms rather than in the abstract.

For the GenAI Appendix specifically, a rubric does something beyond structuring your grading. It restructures student behavior before the appendix is submitted. An ungraded appendix gets treated as a formality because it’s merely a printout of a chat session that students attach without much thought. Grade it on the criteria below, and the appendix becomes part of the assignment. Students who know their prompt quality is being assessed write better prompts. Students who know their verification process will be scrutinized verify their citations. So the rubric incentivizes thoughtful work.

### How to Build the Rubric

I suppose there is a certain irony in using GenAI to build a rubric for evaluating how well students used GenAI.

The following steps walk through the process.

## Step 1: Identify Your Criteria

Before opening any GenAI tool, decide what you are evaluating. For the GenAI Appendix, the criteria map directly to the assignment's purpose.

- Whether the prompts in the appendix are specific and purposefully constructed, not a restatement of the assignment description
- Whether the student read the GenAI output critically and identified errors or gaps rather than accepting it wholesale
- Whether the final work product shows substantive revision, not cosmetic editing
- Whether citations were independently verified against primary sources rather than carried over from the GenAI output unchecked

You may want to add or substitute criteria based on your course. An upper-level seminar might weight critical evaluation more heavily than prompt design. A first-year writing course might add a criterion for the quality of the student's reflective narrative. Adjust before you start prompting.

## Step 2: Choose Your Scale

The rubric below uses a 4-point scale (4 = Excellent, 3 = Proficient, 2 = Developing, 1 = Inadequate) applied to four criteria, for a maximum of 16 points. If your course uses a different grading structure, divide the student's score by 16 and multiply by whatever point value you assign to the appendix.

## Step 3: Prompt the GenAI

Here is the prompt used to generate the first draft of the rubric at the end of this appendix. The structure of the prompt illustrates what Step 1 is asking you to do.

### Sample Prompt

You are a law professor building a grading rubric for a GenAI Appendix that students submit alongside legal writing assignments. The appendix should demonstrate four things: (1) specific, well-designed prompts rather than generic requests; (2) critical evaluation of GenAI output, including identification of errors or gaps; (3) substantial human revision and verification, not light editing; and (4) proper citation checking against primary sources.

Create a rubric using a 4-point scale: 4 = Excellent, 3 = Proficient, 2 = Developing, 1 = Inadequate. Write full performance-level descriptors for each criterion at each score level. Descriptors should be concrete enough that two faculty members grading the same submission would reach the same score — meaning they should describe what appears in the submission, not just how well the student performed. Format as a table. Plain language throughout; no corporate training-manual phrasing.

Notice what the prompt includes: role, task context, specific criteria, the scale with labeled levels, explicit guidance on descriptor quality (concrete enough for two graders to agree), format preference, and a tone instruction. That last element matters more than it might appear. Without it, you tend to get descriptors that read like a compliance checklist, which tells you nothing useful when you are working through a stack of submissions.

## Step 4: Review the Output

GenAI will produce a serviceable first draft, but be aware of these patterns when reviewing it:

**Blurring between the 3 and 2 levels.** The descriptors for Proficient and Developing often look nearly identical, particularly for criteria that involve judgment. If you cannot tell from the rubric what separates the two levels, your students cannot either. Push back on the GenAI directly: *The descriptors for Proficient and Developing under Criterion 2 are too similar. Make the distinction more concrete — what specific difference in the submission separates these two levels?*

**Abstract language.** Phrases like “demonstrates an awareness of” or “shows understanding of” are almost impossible to apply consistently across a set of submissions. Replace them with observable content: what appears in the appendix, or what is absent from it.

**Missing course context.** A rubric written for a contracts memo will need adjustment for a criminal procedure research assignment or an administrative law brief. The GenAI does not know your course.

## Step 5: Revise and Test

Before finalizing the rubric, apply it to a sample anonymized submission such as a strong submission from a prior semester or one you have already graded. If your rubric score does not match your assessment of the work, figure out why. Either the rubric is missing something or your intuitive grading has been inconsistent. Both are fixable, but you want to find out before the rubric goes to students.

If a colleague is willing to score the same submission independently, that calibration exercise is worth the thirty minutes. Disagreements between two graders reveal exactly which descriptors need more precision.

## The Rubric

### GenAI Appendix Evaluation Rubric — Total: 16 Points (4 criteria × 4 points each)

Criterion	4 — Excellent	3 — Proficient	2 — Developing	1 — Inadequate
<b>1. Prompt Quality</b>	Prompts are specific, context-rich, and purposefully structured. The appendix shows that the student assigned a role, described the task in concrete terms, specified format or constraint requirements, and used multiple iterations to refine the output. Prompt choices reflect command of the subject matter — not a cut-and-paste of the assignment description.	Most prompts include meaningful context and clear direction. Some iteration is evident, and prompts appear to have been refined in response to earlier output. A few prompts may be generic, but they do not define the overall approach.	Prompts are broad or formulaic — recognizable versions of the assignment description rather than purpose-built instructions. Little or no iteration is visible. The same prompts could have been submitted for almost any legal writing assignment.	Prompts are one-line generic requests (e.g., "Write a memo about Terry v. Ohio") with no context, no role assignment, and no constraints. No evidence of iteration or deliberate refinement.
<b>2. Critical Evaluation of Output</b>	The narrative identifies specific errors, hallucinations, or analytical gaps in the GenAI output and explains why each is a problem. The evaluation reflects independent	The narrative identifies the most significant problems in the output and offers substantive commentary. The evaluation is meaningful, though it may miss secondary issues or	The narrative acknowledges GenAI limitations in general terms ("AI can make mistakes," "I checked for accuracy") but does not identify specific errors or explain what was wrong	No meaningful evaluation of the GenAI output appears in the appendix. The student either accepted the output without question or offered only a sentence of vague

Criterion	4 — Excellent	3 — Proficient	2 — Developing	1 — Inadequate
	legal knowledge — the submission identifies not just that the output was wrong, but where and why. Commentary goes beyond generic observations about AI limitations.	lack full explanation for why particular output was flawed.	with particular portions of the output. Reads as a pro forma disclaimer rather than genuine engagement.	disclaimer. No evidence the output was read critically.
<b>3. Human Revision and Verification</b>	The final work product is substantially different from the GenAI output in the appendix. The narrative explains specific revisions made and the reasoning behind them. Changes reflect genuine legal analysis, not cosmetic editing. The submission demonstrates clear faculty ownership of the final product.	Meaningful revisions are evident and the narrative addresses the most significant changes. Real work beyond light editing is apparent, even if some sections of the final product closely track the GenAI output — with explanation for why those portions were retained.	Revisions are minor or cosmetic — sentence-level edits, formatting changes, small additions. The final work product largely tracks the GenAI output. The narrative describes revision but the submission itself does not support that description.	The final work product appears to be the GenAI output with minimal or no modification. Any narrative describing verification steps is not reflected in the submission itself.
<b>4. Citation Verification</b>	All citations in the final work product have been independently verified against primary sources. Errors, hallucinated citations, or misrepresentations found in the GenAI output are documented and corrected in the narrative. The verification process is described with enough specificity to assess whether it was rigorous.	Most citations are verified and identifiable errors corrected. The verification process is described in the narrative. Minor gaps may exist, but the overall approach is methodical rather than selective.	Citation verification is described but appears incomplete. Some citations in the final work product are unverified, or errors from the GenAI output carry through uncorrected. The narrative mentions verification without demonstrating it.	No evidence of independent citation verification. Citations appear to have been reproduced from the GenAI output without checking. Hallucinated or inaccurate citations appear uncorrected in the final work product.

## Score Conversion

If you need to convert the 16-point scale to a percentage or letter grade, divide the raw score by 16. The table below provides a rough guide, though your institution's grading scale governs.

Raw Score	Percentage	Approx. Grade	What It Signals
16	100%	A	All four criteria at Excellent
14–15	88–94%	A–/B+	Strong work; minor gaps in one or two criteria
12–13	75–81%	B/B–	Competent but uneven; meaningful gaps in at least one criterion
10–11	63–69%	C+/C	Developing; significant gaps in multiple criteria
Below 10	Below 63%	D or lower	Substantial deficiencies across criteria

## A Note on Weighting

The rubric weights each criterion equally, which is a reasonable default. If citation verification matters more in your course than prompt design, or if your students are just beginning to learn how to prompt and you want that criterion weighted differently, adjust the point values before distributing the rubric. Whatever you decide, make it explicit. Students whose grading criteria are transparent before they submit produce better work than students who encounter the rubric for the first time at the bottom of a returned assignment.

## Sharing the Rubric with Students

Distribute the rubric when you assign the GenAI Appendix, not when you return grades. Students who see the evaluation criteria in advance have a genuine opportunity to meet the standard before submitting. Students who see the rubric for the first time on a graded paper mostly just feel blindsided and want to argue about it. Given that the whole point of the appendix is to develop sound habits around GenAI use, you want students engaging with these criteria while they are working, not after the fact. The rubric is a teaching tool as much as a grading instrument.