

Rubric Design Guidelines SECOND DRAFT (see the course PPT on Performance Assessment)

Summarized *mostly* from Popham, W. J. (2014). Classroom assessment: What teachers need to know (8th edition). Pearson.
Compiled by mjstuve@bsu.edu for my students in EDPS 345, 393, etc. • Last update Dec. 1, 2025

ABOUT RUBRICS

As the measurement instrument in performance assessment, a rubric specifies the rules and metrics to evaluate quality of a student's performance or product:

- The **criteria** or domains of performance are the **rows**
- The **performance levels** are the **columns**
- The gradations or indicators of quality per criteria are the **cells**
- Rubrics can be used formatively and/or summatively
- Rubrics can be used for self/peer-assessment
- The row (criteria) work for the learning objectives and can have multiple alignments to them and, subsequently, standards.

GUIDELINES FOR PERFORMANCE TASKS (Popham + Wiggins)

1. Generalizable and generative
2. Authenticity
3. Multi-dimensional (foci)
4. A result of instruction (teachability)
5. Fair and Feasible
6. Tasks are significant and complex
7. Tasks are realistic and feasible
8. Tasks require student judgment and self-regulation
9. Tasks typically address multiple, concurrent domains of learning
10. Tasks generate evidence (artifacts of learning) necessary for inferences to be made on objectives and standards
11. Allows opportunities for more advanced student work:
 - 11.1. Rehearsal and practice
 - 11.2. Consulting resources
 - 11.3. Citing evidence
 - 11.4. Making claims
 - 11.5. More sophisticated writing
 - 11.6. Processing feedback
 - 11.7. Refining performance to criteria
 - 11.8. Responding to real audience

A related instrument is the task analysis which measures learner independence (autonomy) on a task, whereas a rubric measures the quality.

THE GENERAL PROCESS FOR DESIGNING A RUBRIC

1. Determine the outcomes sought from the task
 - Reference learning objectives and standards
2. Sketch row titles for the criteria (refine over time).
3. Decide on ordinal (column) levels/labels
 - Use contextually common scale when possible, but also resonant (related; supportive) to the task
 - Number of columns matters, as do the perceived gaps
4. Write row descriptions in neutral language
 - Try not to define quality here. Save that for the cells
5. Write a "target" column's cells first
 - Not necessarily the highest column
 - Write as bullet list first, then expand to sentences
6. Paste in other columns and adjust cell text relative to ordinal statements in column levels
7. **Confirm row independence and cell progression**

GOLDEN RULES RUBRIC EVALUATION:

- A. **Criteria (rows)** are **independent** from each other and serve the performance task (learning objectives)
 - B. **Columns** have clear, **distinct** and ordered **performance levels**.
 - C. **Indicators (cells)** are clear and **progressive** relative to the ordinal (column) scale
-

ROW GUIDELINES (Independence):

1. Instructionally addressable (as opposed to innate skills)
2. As few as necessary
3. **Independent from each other**
4. Measurable (via the cell indicators)
5. Serve the performance task and the learning objectives

COLUMN GUIDELINES (Quality):

1. **Ordinal** performance levels set the tone of the outcome
2. Include scoring values that "fit" the ordinal AND customs/conventions in your classroom or context
3. Order from low to high for easier data analysis

CELL GUIDELINES (Progression)

Goal: Efficiently communicate (indicate) the criteria (row) relative to the ordinal (column) scale definitions

1. Qualified adjustments to an indicator:
 - Proficient: "Writing reflects thorough understanding of the material"
 - Basic: "Writing reflects adequate understanding of the material"
2. Quantity of "behaviors" required within an indicator:
 - i.e. "No more than three typos"
3. Conditions of multiple subindicators (e.g. use of AND and OR to denote performance level):
 - i.e. "Demonstrate at least two of the following three techniques in performance..."

TRIAGE FOR RUBRIC REPAIR

1. Problems of metrics (e.g. incongruent scales, etc.)
2. One rubric should really be two or more (or clustered criteria rows)
3. One row should really be two or more (i.e. there are multiple constructs in a criteria)
4. A row is not needed at all
5. Cell text not sensitive to real performance (on criteria)
6. Cell text is too sensitive
7. Criteria (row descriptions) not written in neutral language
8. Columns labels (ordinal scale) aren't differentiated
9. Column straddling: Cells are technically differentiated, but student performance blurs across columns. Solution: Re-qualify your indicators across columns better
10. Untaught domains (akin to Popham's teachability)