Good tasks have the following features, dimensions, or traits:

- *Content*—The task elicits the correct performance from the student.
- *Clarity*—Students know exactly what to do.
- *Feasibility*—The task is practical.
- *Fairness and Accuracy*—Nothing in the task will give an inaccurate picture of student skill. All students have an equal chance to achieve.
- *Sampling*—The task(s) adequately cover all dimensions of the learning target to be assessed.

Content

The job of the task in a performance assessment is to elicit the right performance from students so that proficiency can be accurately judged using the rubric. Thus, the performance task, the rubric, and the student learning targets to be assessed must match up—they must align.

Please notice the difference between our notion of sound content in a performance task and those in certain other task development guidelines. Others tend to focus on "authenticity" and use such terminology as *open, active, work with other students, fosters persistence, challenging,* and *rich.* While these might be relevant for many of the tasks we design to assess complex targets, they are not criteria for all performance tasks. It is perfectly possible to have simpler performance tasks that assess simpler targets. Consider, for example, reading rate. The task is relatively simple—ask students to read aloud and calculate the number of words read per minute.

Our rule of thumb is this: Simple target, simple task; complex target, complex task.

The important thing is that the task elicits the intended performance so that it can be observed and evaluated. Since many targets evaluated with a performance assessment are complex, performance tasks tend to be complex, open, active, and challenging. However, they don't have to have all these characteristics to be sound.

Other Traits

Many things can go wrong with a task. Even if the content, in concept, is perfect for the job, in execution the task might have problems. For example, asking students to work in groups to build a mousetrap car to assess student understanding of mechanics, problem solving, scientific method, group process skills, and communication skills, might be a good idea—the *content* is sound. But, the actual instructions might not be clear (*clarity*), or the resources or time might not be equally available to all students (*fairness*), or it might be hard to judge individual skills in the context of group work (*accuracy*). A good task avoids these sources of bias and distortion.

Interpreting the Scale Points

The descriptors under each level of performance in the rubrics are not meant as checklists—not everything listed must be present to rate a "Ready to Use." Rather, the descriptors are meant as indicators that help the user focus on the level of quality.

The descriptors under each trait in the Metarubric are not meant as a checklist. Rather, they are indicators to help the user focus on the correct level of quality of the rubric under consideration. A strong score ("Ready to Use") doesn't necessarily mean that the rubric is perfect; rather, it means that you would have very little work to do to get it ready to use. A middle-level score ("Needs Refinement") means that the rubric is strong in some ways but weak in others—it would take some work to make it usable, but it is probably worth the effort. A weak score ("Needs Major Revision") means that the rubric needs so much work that it is probably *not* worth the effort—it is time to find another one or begin from scratch.

Please note that a middle score does not mean "average"; this is a criterion-reference scale, not a norm-referenced one. It is meant to describe levels of quality in a rubric, not to describe what is currently available. The "average" rubric currently available may, for example, in fact be closer to "weak" than to a "middle."

The scale could easily expand to a five-point scale. In this case, think of "4" as a balance of characteristics from the "5" and "3" levels, and a "2" as a balance of the "3" and "1" levels.

Summary

Content

- Elicits the right performance.
- Aligns with standards/targets and performance criteria.
- Simple target ⇆ Simple task
  Complex target ⇆Complex task

Clarity

- Students know exactly what to do.
- Students are reminded of the performance criteria.

Feasibility

- The task is practical.
- Students have enough time.
- Proper materials and equipment are available.
- Rating can be accomplished within the time allowed.

Fairness and Accuracy

- Nothing in the task will give an inaccurate picture of student skill.

Sampling

- The task(s) adequately cover all dimensions of the learning target to be assessed.
- Just enough tasks to show students' level of proficiency.

Content

The content of an exercise or task constitutes what the student is asked to do.

Ready to Use

- The task **fits the target**(s) to be assessed; it will likely elicit the desired skills or products. Simple target ⇆ simple task; complex target ⇆ complex task.
- The content of the task represents **important**, core aspects of the curriculum.
- The task is **worth the time** spent on it. It is an episode of learning.
- The task **models good instruction**.
- The task clearly **communicates appropriate** standards and **expectations** to students—what is important to be able to do, and how such knowledge and skills are to be communicated.
- The amount of **scaffolding** in the task **fits the target** to be assessed.
- The task is **developmentally appropriate**.
- The task will **motivate students** to do their best. It is engaging.

Needs Refinement

- There is some question about whether the task matches the target, is important and worth the time spent on it, and so on.
- The task fits the target(s) to be assessed, but there are aspects that might not be engaging for students.

Needs Major Revision

- *Fatal problem:* The task doesn't seem to fit the target(s) to be assessed. It would probably not elicit the desired skills or products.
- You find yourself asking, "Why spend time on this? Is this task worth the time devoted to it?"
- The task could give students a skewed picture of what is important to know and be able to do.
- The task could give students a skewed picture of the role of students in assessment.
- Scaffolding does *not* seem to fit the target to be assessed.
- The task is too hard or too easy for students; it is not developmentally appropriate.
- Students will probably not be motivated to do their best.
- The task seems to be overkill for the target to be assessed—a simpler task would be better.

Clarity

How clear are the instructions? The content may be good, but there may be some unintended ambiguity in what students are asked to do.

Ready to Use

- Instructions are **clear and unambiguous**.
- The achievement expectation(s)—target(s)—are **clearly stated**.
- The **specific** kind(s) of performance or **work to be demonstrated is stated**.
- The **context and conditions** within which proficiency is to be demonstrated **are detailed**—time limits, materials available, location, resources that can and can't be used, and so on.
- Students are reminded of the criteria by which their work or performance will be judged. These **performance criteria match the task**.
- Any **ambiguities in the instructions are intentional**—the goal is to see the extent to which students can figure out what they should do.

Needs Refinement

- The exercise or task is more or less clear, but could be improved. The body of the activity is clear; all that remains is details.
- The task seems clear enough, but students are not reminded of the performance criteria.

Needs Major Revision

- The instructions are not clear.
- Terms are unfamiliar or not well defined.
- Students are not reminded of the criteria by which their work or performance will be judged, or these reminders are incomplete or inaccurate.
- You find yourself asking, "What do they mean by this? I could respond to this question in more than one way depending on how I interpret the instructions."
- Ambiguities are unintentional.
- The task has so many problems that it might be better to start over.

Feasibility

"Best" has to be balanced against "practical."

Ready to Use

- Students have **enough time and resources** to complete the task to the best of their ability.
- The task is **safe**.
- **Materials and equipment are available** and can be packaged quickly, as needed.
- If **compromises** have been made between feasibility and "best," there is a **clear and reasonable** description of the tradeoffs and why the final design was decided on. The task will still elicit the intended knowledge, product or performance by students.
- If **1:1** observation is required, it is **handled efficiently**, with lots of assistance from trained observers.

Needs Refinement

- Lots of 1:1 observation is required, and, although plans are in place for adequate observers, logistics are incomplete.
- Some aspects of the design might not be feasible, but adequate refinements are possible.

Needs Major Revision

- The task is not safe.
- The task will take too long.
- Needed resources (personnel, expertise, equipment, materials) probably would not be available. So many compromises have been made between "feasible" and "best" that the task might no longer elicit the performance intended.

Fairness and Accuracy

Does the task require any skills or knowledge, unrelated to the target(s) being assessed, that might cause students not to do their best?

Ready to Use

- The task **avoids potential sources of bias and distortion** (see examples under "Needs Major Revision") that might result in an inaccurate picture of student ability.
- The task **allows for individual learning styles and needs**. This can include accommodations, but only if they are unlikely to alter what is intended to be measured.
- The **task statement is fair**. There is nothing in the way in which the task is stated that would cause any group of students to be unable to demonstrate fully what they know and can do. The task is not dependent on unique background knowledge, individual circumstances or student characteristics that would enable one group to do better than another for reasons unrelated to the knowledge or skills being assessed.
- Student **response requirements are fair**. There is nothing in the way in which students must provide their response to the task that would cause any group of students to be unable to demonstrate fully what they know and can do.

Needs Refinement

- The task could benefit from other response alternatives; it is easy to see how the task could be amended.
- Although some aspects of task content might be unfamiliar to certain groups of students, it is easy to see how the task could be revised.
- It is easy to see how accommodations, although missing, could be added.

Needs Major Revision

- There are things about the task that would probably lead to an inaccurate picture of student ability on the target(s) to be assessed. For example, there might be too much reading or writing on a math test; students might be asked to demonstrate skill in front of an audience when private performance would work just as well, and so on.
- Not all students will have equal access to the materials and resources needed to do their best on the task.
- Some *groups of students* might be systemically unable to do their best on the task due to task features that have nothing to do with the knowledge or skill being assessed—task context, response format, working conditions, requirements for assertiveness, requirements for public demonstrations, and so on.

Sampling

Sampling refers to how well a task (or set of tasks) represents the breadth and depth of the achievement target to be assessed. The major question to answer is this: "If a student does well (or poorly) on this task (or set of tasks), can I make appropriate inferences regarding the student's overall skill level?"

Ready to Use

- The **target product or skill is defined well** enough that one can judge the degree to which a given task or set of tasks is representative.
- When students are allowed to make choices on which task to complete, such **choice will not bias results**.
- If you are working at a group of **tasks**, they **sample(s) appropriately** from the possible breadth and depth of the skill or body of knowledge to be assessed.
- If you are looking at a single task, it is easy to see how a single task **fits into the "big picture"** of assessing overall skill.

Needs Refinement

- It seems like the given set of tasks is representative, but the target product or skill needs to be a little better defined to be sure.
- The task represents an important part of the achievement to be measured, but explanation is required of what is left out and why.
- The task or set of tasks represents a somewhat restrictive view of the target, but it is fairly easy to see what could be amended.

Needs Major Revision

- It is difficult to know how well the task samples the target because the target is not very well defined to begin with.
- The task(s) represent(s) only a very small portion of the achievement target.
- You find yourself asking, "Why did they choose to assess that target using this task? It represents a pretty narrow definition of the target."