USAID
FROM THE AMERICAN PEOPLE

MECap Knowledge Product

Rubrics for Evaluation and Learning

By Heather Britt, Senior M&E Specialist, PPL/LER/MECap

A rubric is a guide for assessing performance by defining quality and value. Rubrics are particularly helpful for synthesizing data from multiple sources to provide a credible answer to evaluation or learning agenda questions.

Overview of Rubrics

A rubric is a tool for making well-reasoned and well-evidenced judgements of merit, worth and significance.

Rubrics guide judgement by describing how quality and value are defined. A rubric specifies criteria that define a quality being judged. A rubric may be designed for a single criterion or multiple criteria. For example, the Inter-Agency Standing Committee (IASC) for humanitarian aid promotes use of its Gender Marker to assesses interventions against a single criterion, gender equality. The Multimedia Project Evaluation Rubric contains eleven criteria related to the quality of multimedia. For each criterion levels of performance describe how well what is being evaluated (the evaluand) performs. The levels of performance are ordered from lesser to greater (or vice versa) according to the degree that they manifest the criterion[i]. Performance levels may be assigned either a quantitative or qualitative label (such as poor, average, excellent). Both the IASC Gender Marker and the Multimedia Project Evaluation Rubric assign numbers to each performance level.

Rubric Examples

The IASC Gender Marker is a tool for coding whether or not a humanitarian intervention is designed well enough to ensure that women/girls and men/boys will benefit equally from it or that it will advance gender equality in another way.

The Multimedia Project Evaluation Rubric contains eleven criteria and four grading options for assessing multimedia projects.

Similar but not the same

When deciding whether to use a rubric, consider these two similar, but different, measurement tools:

- Checklist: Use a checklist when it is enough to verify the presence or absence of criteria. Descriptions of performance quality and levels of performance are minimal or lacking.

- Interval or ratio scale: Use an interval or ratio scale when levels of performance can be described precisely as constant units of measurement spaced in equal intervals. A ratio scale has a clear definition of zero. A rubric is an ordinal scale – the levels of performance are arranged in order but the distance between them cannot be specified.

Using a rubric

To use a rubric, those conducting a learning agenda inquiry or evaluation would review the relevant evidence, compare the evidence to the descriptions of each performance level, and make a judgement. One common example of rubric use is student performance grading. The teacher reads a student essay, consults an assessment tool, and assigns a grade of A, B, C, or D.

The evidence used to make an assessment may be qualitative or quantitative, and come from one source or many. To use a rubric such as the IASC Gender Marker rubric in an evaluation, an evaluator may review documents, interview staff and beneficiaries, and conduct observations of the intervention in action. Triangulating data from reliable sources increases the likelihood that judgements will be viewed as credible.

Absolute or relative judgements?

Rubrics are appropriate for judging both absolute or relative merit and worth. Absolute judgements involve grading an evaluand against a recognized standard. The minimum acceptable level of performance on a criterion is referred to as a bar.

Examples of questions of absolute merit that could be answered using a rubric in the USAID context:

- Does this Activity MEL Plan comply with the Automated Directives System (ADS)?
- Is this activity likely to contribute to gender equality (as defined by the IASC Gender Marker)?
- Does this intervention meet the Core Humanitarian Standard on Quality and Accountability?

Relative judgements of quality are made in comparison to one or more other evaluands. Relative judgements involve ranking the evaluands under consideration. Examples of questions of relative merit that could be answered using a rubric:

- Which of the Mission's five Activity MEL Plans is likely to provide more rigorous data?
- Which of our three pilot interventions is most likely to contribute to gender equality?

When to Use a Rubric?

Rubrics provide credible answers to evaluation and learning questions. They are a flexible tool that can add value to several types of inquiry including learning agendas, evaluations, special studies, and evaluation synthesis, also referred to as meta-evaluations. Rubrics are particularly helpful for dealing with large amounts of data from multiple sources.

Rubrics are most useful when developed prior to data collection. However, they can also be introduced later when preparing for analysis. When used before data collection, rubrics facilitate agreement on the type and amount of data that evaluation or learning agenda users will find credible.

Such an agreement may reduce data collection or eliminate unused data. When used to guide analysis, they may streamline the process by specifying what to look for when examining the evidence.

Finally, rubrics may increase the usefulness of evaluations and studies by synthesizing findings in a transparent way. Rubrics can produce a single answer to an evaluation or learning question, rather than a number of findings that require further consideration. Evaluation users can be confident in that answer because rubrics describe how evidence has been used to make a judgement.

Creating a Rubric

Theoretically, creating a rubric involves the following four steps. In practice, the process is likely to be iterative.

1. Select criteria to include in the rubric
2. Organize the criteria within the rubric
3. Outline performance levels for each criterion
4. Describe performance at each level

Ideally, the intended users of a rubric should participate in its design. To design a rubric, stakeholders discuss and agree on how ‘quality’ and ‘value’ is defined in their work. The collaborative process helps ensure that the rubric will be credible and useful. And it is a great way to build the evaluative thinking needed to generate, understand, accept, and use evaluation findings.

Select criteria: Start wide, end narrow

Rubric designers may find it helpful to start wide and end narrow, that is, begin the process by considering a wide-range of possible criteria before making the final selection. To start wide, brainstorm important aspects of the desired quality long enough to get past the conventional ideas and uncover elements essential to a useful assessment. Those who will use the evaluation or learning agenda study can help identify criteria and clarify their relative importance. Consider consulting with those with practical experience of the evaluand, as well as the literature.

When it is time to winnow the list of possible criteria, start by consolidating concepts that are very similar. A small pilot or a simple thought experiment can inform the selection process. Consider real cases of both positive and negative experiences of the desired quality.

Strategies for selecting criteria for a rubric[1]

Strategy	Advantages and disadvantages
Stakeholders vote	Inclusive and democratic, but assumes that stakeholders are all well informed.
Selected stakeholders provide input; assessor makes final selection	Combines stakeholder and assessor expertise, but requires careful justification.
Use evidence from the literature	Avoids reinventing the wheel and is a good supplement to stakeholder input, but requires good knowledge of the literature.
Use specialist judgement (expert panels)	Quicker than a literature search, but may reflect only the prevailing expert view.

When using a rubric as part of an evaluation, criteria may be derived from needs assessment reports, the theory of change, or design documents. Not all stakeholders may be familiar with these sources and how they inform criteria selection.

Organize the criteria within the rubric

Consider whether the final judgement should be made holistically, or by considering each criterion individually. Consider whether criteria should include a bar, or minimum performance level. Consider whether the criteria are equally important, or they should be weighted differently. For example, do we need qualifying criteria, "bonus points?"

Considering criteria individually or holistically

There are many different ways to design rubrics to meet users' needs and organize criteria within a rubric. For example, the rubric may specify that each criterion must be met. In that case, an evaluand must meet the minimum performance for each criterion to avoid a failing grade. If the evaluand does not meet the bar for a single criterion, its overall performance is considered substandard.

Another approach involves grading the evaluand on each criterion, then coming to the final judgement by considering the performance grades holistically. In this approach, the rubric as a whole informs the final judgement.

A third approach involves weighting each criterion differently to indicate its importance in the overall score. For example, one or more criteria may be used to determine which candidates qualify for assessment. Only an evaluand that meets the bar of the inclusion criteria would be assessed with main rubric. One or more criteria may provide "bonus points" in a holistic assessment. Such criteria are considered "nice to have" rather than essential criteria.

The rubric's layout or format should also aid assessment. For example, two or more criteria may be arranged to create a matrix. When making a complex assessment, or synthesizing large amounts of data, rubrics may be "stacked" in such a way as to reach a single judgement.

[1] Adapted from Davidson, E. J. (2005). Evaluation methodology basics: The nuts and bolts of sound evaluation. Thousand Oaks: Sage. (p. 126-127)

Outline the performance levels

Perhaps the first question is: How many performance levels are necessary? When determining the range of performance, do not make the top of the range ideal or unattainable.

What will we call each performance level? Determine whether performance levels will be assigned a quantitative or qualitative label.

Describe performance at each level

When describing performance at each level, include real-life details of what performance looks and feels like. Stakeholders can help provide specific details.

Allow for the variation naturally occurring in the assessment situation. Describe the best performance based on context. List the core elements, but do not be too restrictive.
When describing the top level of performance, use language such as "Virtually all…" or "With no important exceptions…"

When describing the middle levels of performance, you may have trouble making distinctions between them. There are two solutions: 1) reduce the number of levels; 2) learn more about the content and context and revise the levels using with more detailed information.

Using a Rubric for an Evaluation or Learning Agenda

Rubrics are most useful when developed prior to data collection. However, they can be introduced when compiling data in preparation for analysis. This section describes the role of rubrics in guiding decisions about 1) selecting the type and amount of data, and 2) conducting analysis to make a judgement.

Regardless of when a rubric is introduced during an evaluation or assessment, the following six steps describe the process for using evidence to make a judgement.

1. Identify evidence to inform assessment
2. Ensure that evidence will be credible to evaluation users
3. Collect or compile data
4. Analyze data
5. Make judgements
6. Synthesize findings

1. Identify evidence to inform assessments

   Review or create a catalog of all data that is currently available or will be collected. Each criterion of the rubric will probably require different data to make an assessment. Review the type and amount of data that is or will be available and ensure adequate coverage for each criterion.

If additional evidence is needed, consider returning to the sources that informed the criteria selection process and the development of the performance levels (e.g. literature, experts, and key informants).

2. Ensure that evidence will be credible to evaluation users

Provide the evaluation users with a summary of the evidence that will be used to make judgements on each criterion. Ensure that evaluation users find the type and amount of evidence to be credible. A "credibility test" will help determine if additional or different data is needed.

As needed, revise the data collection plan to cover all criteria. Alternatively, revise the rubric to match the available data.

3. Collect data

Collect or compile the evidence needed to make credible judgements on each rubric criterion.

4. Analyze data

Rubrics do not prescribe or limit the type of data analysis, however, the analysis process must include assigning data to each relevant criterion of the rubric. If the data collection plan includes a description of data needed for each criterion, this step is much easier.

Data can be assigned early or later in the analysis process. When working with large amounts of data, or triangulating across several data collection methods, data may be assigned to individual rubric criterion late in the analysis process.

[i] For this reason, rubrics may be considered an ordinal level of measurement.