What should go into an evaluation report rating scale?

As part of a project with an Australian state government agency, I am developing a rubric for people with little to no evaluation skills who might need to judge the quality of an evaluation report. This is within the context of a larger project whereby an evidence base of past evaluation reports is made available for program designers. We want users to access these reports, but also have some support to judge the quality of the reports.

I have searched a number of existing tools and assembled an (over-inclusive) list of items that could be part of an evaluation report quality rubric. Items are grouped around common themes that existing rubrics use. There are also some comments that help explain each item. It is anticipated that each item in the rubric will be rated on a simple No/Partially/Yes/Not applicable scale.

Feedback wanted:

The questions I have that I would like feedback upon, are:

1.   Are these the appropriate categories for organising the items?

2.   Are the items appropriately worded?

3.   Are some unnecessary?

4.   Are some missing?

5.   What other comments are needed to help explain the items?

About this rubric:

Who is the audience?

The rubric is being designed for users that will include people who have very limited evaluation experience. This will include people designing programs within the relevant policy space.

What is the aim of the rubric?

We want users to be able to be critical consumers of evaluation reports. This rubric aims to provide users with a minimum set of criteria to judge the quality of evaluation reports.

How will this rubric be used?

This rubric will be made available alongside an online evidence database containing a range of evaluation reports. It is hoped that users of the database will be able to use the checklist to assess the quality of the reports in the database, rather than simply drawing upon these reports without thinking critically about the quality.

Quality rubric for evaluation reports

Items	Comment
Presentation	
Were the objectives and scope of the evaluation clear and reasonable?	
Is the evaluation written in clear and plain English?	No technical matter, formulas, or excessive graphs and tables in body of report
Is the program or project being evaluated clearly described?	Main components of the program are described; timeline for delivery of program; possibly included in the program logic
Is the evaluation organized around KEQs or recommendations?	Should not be organized by methodology; analysis should not be separate from interpretation
Design	
Was the evaluation guided by a clear theory of change?	A narrative explanation about why the program is believed to work
Was the evaluation guided by clear outcomes and program logic?	Only partial if it uses a pipeline logic model without causal sequence
Is the evaluation guided by clear and appropriate Key Evaluation Questions?	
Was the research design suitable for answering the KEQs?	
Did the evaluation draw on the evidence base?	Explains how past evidence was found; how it was used to inform methodologies and/or to compare findings
Have clear criteria been set for judging the 'value' of the program?	Makes clear what form the final outcome(s) should have for the program to be a success
Was the evaluation ethically informed?	Use of ethics approval process; informed consent from participants; culturally sensitive where appropriate
Methodology	
Are the sources of data clearly described and aligned to the KEQs?	For secondary data clear link to sources and discuses validity For primary data key parts of the methodology such as survey instrument, interview schedules, are transparent
Is the method of sampling clearly described and aligned to the KEQs?	Identifies area and population to be represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects, limitations to sample
Are important relationships assessed through multiple methods where appropriate?	For example, uses both subjective and objective measures for key variables; or different forms of data collection
Did the evaluation look for perverse/contradictory findings?	
Are methodological limitations and their impact on the evaluation explicitly discussed?	Discusses possible sources of bias e.g. potential interviewer bias

Analysis	
Was each method of analysis rigorous in terms of recognized standards?	For example, were qualitative data analyzed through a documented coding process.
Does the analysis make appropriate comparisons to address the KEQs?	For example, comparison groups based on types of program intervention or by demographic characteristics
Were external factors adequately accounted for in the analysis?	
Interpretation, findings, and recommendations	
Was statistical significance interpreted as practical significance?	
Are the findings consistent with the data and analysis?	
Were the limitations of the evaluation explicitly stated and taken into account in findings?	
Is any attribution of program impact justified?	For example, was correlation confused with causation?
Were alternative explanations for results considered?	
Were results compared with those of similar programs?	
Were recommendations clear and actionable?	

Sources:

ACT Government, 2010, *Evaluation Policy and Guidelines*,

UNIFEM, 2009, *Guidance: Quality Criteria for Evaluation Reports*, Evaluation Guidance Note Series, No.8

UK Department for International Development, EQUALS, Quality Assurance Evaluation Report,

European Commission, 2015, *Quality control in the evaluation reports*, https://europa.eu/capacity4dev/evaluation_guidelines/minisite/en-geographic-thematic-and-other-complex-evaluations/synthesis-phase-3/quality-control-eval