
BVEM Evaluation Rubric
Background, Design, and Implementation

BUSINESS RESEARCH DIVISION

Business Research Division
Leeds School of Business
University of Colorado Boulder
420 UCB
Boulder, CO 80309-0420
Telephone: 303.492.3307
[Colorado.edu/business/brd](https://colorado.edu/business/brd)
Contact Person: Brian Lewandowski
Brian.Lewandowski@Colorado.EDU

August 30, 2023



Leeds School of Business
UNIVERSITY OF COLORADO **BOULDER**

I. Purpose of Study

The Business Research Division at the Leeds School of Business was retained to develop and employ a *consistent* and *objective* rating system for the labor component of Public Service Company of Colorado's (PSCo) electric resource acquisition proposals, which aligns with the Best Value Employment Metrics (BVEM) regulation.

II. BVEM Overview

Since 2010, statute has required employment metrics to be a consideration when the Colorado Public Utilities Commission evaluates and approves utilities' electric resource acquisition projects. Specifically, §40-2-129, C.R.S. was revised to state that the utility was required to "obtain and provide" the Public Utilities Commission of Colorado (PUCC) with BVEM information. As a result of this change, the PSCo provided specific guidelines for bidders on BVEM when preparing responses to the 2022 All-Source Solicitation.

BVEM is categorized into four general areas: availability of training programs; employment of Colorado workers as compared to importation of out-of-state workers; long-term career opportunities; and industry-standard wages, health care and pension benefits. The requested information, by craft and job classification, specifically includes:

1. Availability of training programs:
 - Availability of training programs;
 - Names of specific training programs available;
 - Curriculum of specific training programs;
 - Cost of worker training;
 - Duration of training programs;
 - Total number of hours of on-the-job training required;
 - The total number of classroom hours required;
 - The licenses and certifications obtained (if any);
 - Copy of training program standards for each program;
 - Statement whether the training programs are United States Department of Labor registered apprenticeship programs and are accredited to award college credits.
2. Employment of Colorado workers as compared to importation of out-of-state workers (by job classification):
 - Estimated number of workers;
 - Estimated length of time of service, including total man hours;
 - Percentage of Colorado workers;
 - Percentage of project man hours earned by Colorado workers.
3. Long-term career opportunities (by craft and job classification):
 - Licenses, certifications and skills that will be applied;
 - Long-term career opportunities.
4. Industry-standard wages, health care and pension benefits (by craft and job classification):
 - Range of wages;
 - Healthcare benefits;
 - Pension benefits;

- Prevailing wages and fringe benefits based on industry standards and the current Colorado labor agreements;
- Wages and fringe benefits.

III. Literature Review

A significant portion of the research on best practices in designing and implementing rubrics has been published by two groups: the Association to Advance Collegiate Schools of Business (AACSB) and ETS. The AACSB is a nonprofit association who provides quality assurance and accreditation standards for over 1,850 member organizations and more than 950 accredited business schools worldwide.¹ Similarly, ETS is responsible for designing, administering, and scoring AP testing worldwide and explicitly focuses on equitable assessment.² Both agencies rely strongly on reputations of objective and consistent scoring and as such have invested significant resources into understanding how to develop and implement high-quality rubrics. Because of the nature of these groups, the majority of this literature discusses rubrics in the context of education, however, they are also employed in objective scoring/rating in other industries such as policy analysis and healthcare. Objectively, the best practices are consistent across use-cases.

For a “good” functioning rubric, there is a trade-off between generalizing factors to be scored so that it is functional and applicable in an objective way and keeping the rubric specific enough to allow differentiation between proposals. This is also referred to as “holistic” scoring within the literature.³ In some cases, rater cognition is important: the mental process that raters use to arrive at a score. More commonly, rubric outcomes value consistency between different raters, known as inter-reader agreement.

Holistic scoring began as a way to rank or compare candidates and not measure their levels of knowledge. In other words, the index or score from the rubric does not provide information regarding skill or knowledge levels, but provides information on their knowledge *compared* to other test-takers. Today rubrics can be holistic and also provide measures of knowledge (modified holistic).⁴ For the purposes of BVEM, the current rubric is holistic in nature as scores do not convey quality of the proposal itself.

IV. Methodology and Criteria

For the rubric to function as intended, BVEM criteria as written needed to be translated into scores. Following best practices as established in the literature, a rubric was developed based on the four categories as defined in Section III (and BVEM guidelines). To increase consistency and reduce personal subjectivity, scoring is based on components which were 1) stated as specifically relevant within BVEM guidelines and 2) independent to inherent differences in structure of the proposal which impact labor

¹ AACSB homepage accessed here: <https://www.aacsb.edu/about-us>

² ETS homepage accessed here: <https://www.ets.org/about.html>

³ Bejar, Isaac. *Advancing Human Assessment: The Methodological, Psychological and Policy Contributions of ETS*, Chapter 18: A Historical Survey of Research Regarding Constructed-Response Formats.

⁴ Ibid.

needs.⁵ Bidders provided detail in different areas and forms in the proposal (in the narrative, in tables, and in appendices), presenting a challenge to source information for each section of the BVEM rubric. Each category within BVEM guidelines presented its own unique challenges and solutions. One example from each, along with its resolution, are explained below.

1. Availability of training programs
 - Identified whether any offered trainings are United States Department of Labor-registered apprenticeship programs (yes/no scoring) – this is an accepted designation of quality within BVEM metrics and is explicitly requested in the documentation.
2. Employment of Colorado workers as compared to importation of out-of-state workers (by job classification):
 - Included the percentage of Colorado workers to out-of-state workers and not the raw count – this avoids penalizing less labor-intensive proposals.
3. Long-term career opportunities (by craft and job classification):
 - Examined the explicit discussion of growth opportunities for employees within the company.
4. Industry-standard wages, health care and pension benefits (by craft and job classification):
 - Analyzed whether wages are at or above industry standard (score above average, average, or below average) – this avoids personal judgements regarding differences between 2% or 5% above industry standard.
 - Consider comprehensiveness of benefits (e.g., Healthcare and dental vs. healthcare alone) – this avoids some difficulty of comparing details of benefit plans because quality of a plan depends on personal preference and risk/exposure profiles.⁶

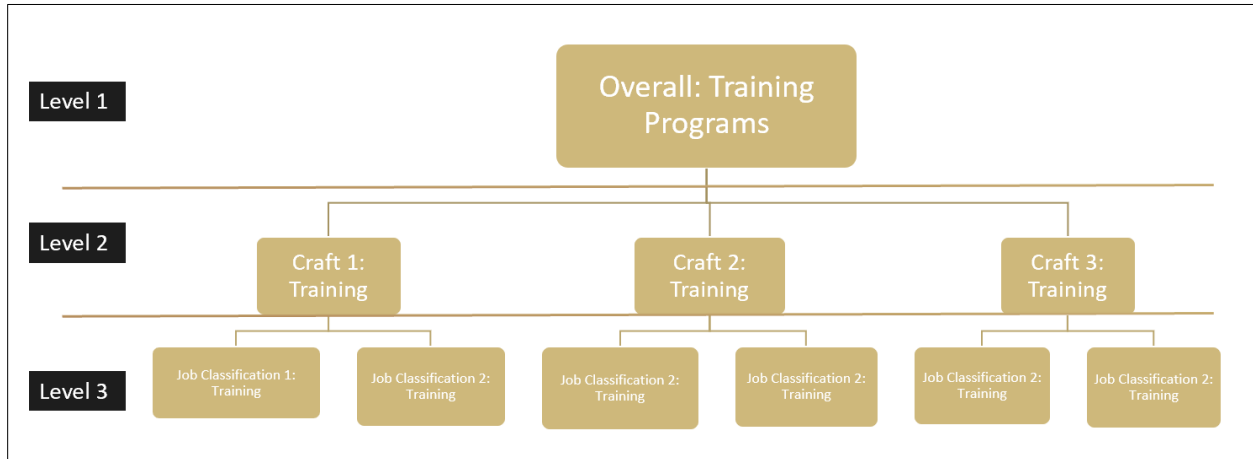
Additionally, given the nature of the proposals, the rubric needed to be flexible to accommodate different levels of detail and different structures based on the type of proposal. Proposals differed by specific type (for example, solar vs. wind) and those differences meant different labor composition and requirements; thus, the intention was to minimize the impact of these differences to the extent that we are able by allowing for bids to be evaluated by resource. One goal of BVEM evaluation was to increase the employment metrics provided with proposals: the scoring structure was designed to increase the incentive to supply detail.

To address these challenges, the BVEM rubric scores responses in what can be described as multi-level buckets. First, bidders earn points by providing general responses in each of the four categories. Second, given the expectation of more detailed responses regarding craft/job classification, bidders were instructed to include details specific to craft and/or job classification within select categories. Figure 1 provides a visual example of the scoring structure.

⁵Also referred to as “reliability” as, for example: Moskal, Barbara M. and Leydens, Jon A. (2000) “Scoring Rubric Development: Validity and Reliability,” *Practical Assessment, Research, and Evaluation*: Vol. 7, Article 10. Available at: <https://scholarworks.umass.edu/pare/vol7/iss1/10>

⁶Dulebohn, J. H., et al., “Employee benefits: Literature review and emerging issues,” *Human Resource Management Review* (2008).

FIGURE 1. VISUAL REPRESENTATION OF BVEM RUBRIC SCORING



This structure incentivized bidders to provide as much detail as possible. A proposal which included detail scored higher than a proposal which did not, regardless of the “quality” of the response. Over time, this will encourage proposals to be more detailed to the point where they are all detailed and quality will be more important.

Scores for each bid and for each component of the rubric ranged between 0 and 1.

Bidders that indicated a project labor agreement (PLA) in place or planned received full credit throughout the BVEM scorecard.

V. Scoring Results

The BRD received 381 bids to score. PSCo reduced the number of bids to 166 for thorough review. Of the 166 reviewed bids, few bids received no points. Recorded scores for each component ranged between 0 and 1, with an average overall score of 0.45, an average general score of 0.57 and an average craft score of 0.38.