

## **Developing a Behavior-Based Rubric for Assessing Professional Skills During Hiring**

**Ms. Nikitha Sambamurthy, Purdue University - West Lafayette**

Nikitha Sambamurthy is a Ph.D. candidate in Engineering Education at Purdue University.

**Dr. Monica Farmer Cox, Purdue University - West Lafayette**

Monica F. Cox, Ph.D., is Professor and Chair in newly created Department of Engineering Education at The Ohio State University. Prior to this appointment, she was a Associate Professor in the School of Engineering Education at Purdue University, the Inaugural Director of the College of Engineering's Leadership Minor, and the Director of the International Institute of Engineering Education Assessment (i2e2a). In 2013, she became founder and owner of STEMinent LLC, a company focused on STEM education assessment and professional development for stakeholders in K-12 education, higher education, and Corporate America. Her research is focused upon the use of mixed methodologies to explore significant research questions in undergraduate, graduate, and professional engineering education, to integrate concepts from higher education and learning science into engineering education, and to develop and disseminate reliable and valid assessment tools for use across the engineering education continuum.

# Developing a Behavior-Based Rubric for Assessing Professional Skills During Hiring

## Abstract

Modern engineers are expected to possess strong leadership, creativity, and communication skills in addition to being technically savvy<sup>[1]</sup>. A recent trend in the engineering hiring process is the inclusion of team-building games used to analyze engineering candidates for professional skills. Current methods of evaluating candidates, such as behavioral interviews, are subjected to bias, with evaluators scoring candidates based on broad and inconsistent definitions of traits, and their own perceptions of how well candidates demonstrate these traits<sup>[2]</sup>.

This methods paper describes a research-informed process for creating a valid and reliable rubric for assessing professional skills in hiring. Step-by-step comparisons are made between rubric development in academic and industry settings, along with a guide for utilizing and analyzing survey responses to inform rubric development, and conducting inter-rater reliability with rubrics. Implications from this guide contribute to further understandings of rubric development in industry.

## Introduction

As globalization increases, today's engineers need to successfully engage in their fields at both the macro level (i.e., working with limited world resources) and micro level (i.e., synergizing in and with engineering teams). Engineers successful in this endeavor must not only be technically savvy, but also possess strong leadership, creativity, and communication traits<sup>[1]</sup>. These skills, along with high ethical standards, dynamism, agility, resilience, and flexibility, have been defined by the National Academy of Engineering as key professional skills successful engineers need.

In response to this requirement, engineering programs have incorporated teaching professional skills through courses, workshops, and incorporating real-world experiences in the classroom through service learning<sup>[3]</sup>. The challenge of teaching of professional skills has been acknowledged by engineering educators, particularly for assessment methods<sup>[4]</sup>. Examples of proposed assessment methods include tests and examinations, measures of attitudes and perceptions (self- and third party reports), portfolios, competency measures (performance appraisals and simulation), behavioral observation, and external examiners; however, there is an inverse correlation between the quality of the assessment and efficiency<sup>[5]</sup>. Shuman et. al.<sup>[4]</sup> additionally identified three major assessment issues for evaluating professional skills in academic settings: (1) Specifying outcome definitions, (2) Understanding the scope of the educational experience needed to master the outcome, and (3) Capturing "awareness" outcomes – students' abilities to understand the importance of the professional skill.

These assessment issues extend to industry, where companies evaluating potential engineering employees are limited to observations and behavioral interviews – and more recently, team-building games – to judge the quality of a candidate's professional skills. These methods are subject to bias due to evaluators scoring candidates on broad and inconsistent

definitions of traits <sup>[2]</sup>, and while academia provides an opportunity for students to identify and improve professional skills, bias in hiring assessments are particularly damning because they directly affect employment. How can we better evaluate professional skills in the engineering hiring process? Can existing processes in academia be translated to industry?

This paper focuses on a research-informed process for creating valid and reliable rubrics for assessing professional skills in hiring. Step-by-step comparisons are made between rubric development in academic and industry settings, along with a guide for utilizing and analyzing survey responses to inform rubric development, and conducting inter-rater reliability with rubrics. Implications from this guide contribute to further understandings of rubric development in industry.

## **Literature Review: Understanding Rubric Development in Academia**

### *What Is a Rubric?*

A rubric is an assessment tool that describes what criteria and expectations are required for a given assignment or task <sup>[6]-[8]</sup>, specifically distinguishing varying quality gradations typically ranging from what constitutes good performance (e.g., an A being earned for scoring 90+ points on a 100 point scale) to poor performance (e.g., a D/F being earned for scoring 60 points or less on a 100 point scale). Within educational environments, rubrics are designed to evaluate a student's process or product <sup>[9]</sup> and have three key features: (1) evaluation criteria, (2) quality definitions, and (3) a scoring strategy <sup>[7]</sup>. A rubric that is co-created with students, used to evaluate peer, self, and teaching assessments, and then used to assign grades is an *instructional rubric*, and a rubric solely used by an instructor to assign grades is a *scoring rubric* <sup>[6]</sup>. Since the goal of this paper is to describe rubric development in industry environments, where instruction is not the goal during the hiring process, scoring rubrics will be the focus of this paper.

### *Different Types of Rubrics*

Rubrics can be classified as general or task-specific (focusing on what aspect of the performance or task to evaluate), and analytic or holistic (focusing on the specific criteria to evaluate for the performance or task), each with its own advantages and disadvantages <sup>[9], [10]</sup>. General rubrics are used to make *generalizations* of performances across tasks with the same learning outcome <sup>[9]</sup>. In these rubrics, criteria refer to facets of the learning outcome instead of task-specific features; for example, listing characteristics of good communication instead of focusing on the content of the speech. In contrast, task-specific rubrics focus on specific tasks and their outcome, such as the specific answer to a math problem and the reasoning they are supposed to use. Analytic rubrics focus on each criterion for evaluating a performance or task outcome, such as each step of a process, while holistic rubrics focus on a performance or task outcome as a whole <sup>[10]</sup>. Table 1 describes the advantages and disadvantages of each type of rubric.

<b>RUBRIC TYPE</b>	<b>DEFINITION</b>	<b>ADVANTAGES</b>	<b>DISADVANTAGES</b>
General	Description of work gives characteristics that apply to a whole family of tasks (e.g., writing, problem solving).	<ul style="list-style-type: none"> <li>• Can share with students, explicitly linking assessment and instruction.</li> <li>• Reuse same rubrics with several tasks or assignments.</li> <li>• Supports learning by helping students see “good work” as bigger than one task.</li> <li>• Supports student self-evaluation.</li> <li>• Students can help construct general rubrics.</li> </ul>	<ul style="list-style-type: none"> <li>• Lower reliability at first than with task-specific rubrics.</li> <li>• Requires practice to apply well.</li> </ul>
Task-specific	Description of work refers to the specific content of a particular task (e.g., gives an answer, specifies a conclusion).	<ul style="list-style-type: none"> <li>• Teachers sometimes say using these makes scoring “easier.”</li> <li>• Requires less time to achieve inter-rater reliability.</li> </ul>	<ul style="list-style-type: none"> <li>• Cannot share with students (would give answers away).</li> <li>• Need to write new rubrics for each task.</li> <li>• For open-ended tasks, good answers not listed in rubrics may be evaluated poorly.</li> </ul>
Holistic	All criteria (dimensions, traits) are evaluated simultaneously.	<ul style="list-style-type: none"> <li>• Sorting is faster than with analytic rubrics.</li> <li>• Requires less time to achieve inter-rater reliability.</li> <li>• Good for summative assessment.</li> </ul>	<ul style="list-style-type: none"> <li>• Single overall score does not communicate information about what to do to improve.</li> <li>• Not good for formative assessment.</li> </ul>
Analytic	Each criterion (dimension, trait) is evaluated separately.	<ul style="list-style-type: none"> <li>• Gives diagnostic information to teacher.</li> <li>• Gives formative feedback to students.</li> <li>• Easier to link to instruction than holistic rubrics.</li> <li>• Good for formative assessment; adaptable for summative assessment; if you need an overall score for grading, you can combine the scores.</li> </ul>	<ul style="list-style-type: none"> <li>• Takes more time to score than holistic rubrics.</li> <li>• Takes more time to achieve inter-rater reliability than with holistic rubrics.</li> </ul>

Table 1. Rubric Types with Advantages and Disadvantages. Modified From Assessment and Grading in Classrooms (p. 201), by Susan M. Brookhart and Anthony J. Nitko, 2008, Upper Saddle River, NJ: Pearson Education.

Rubric developers must first determine whether they will use a general or task-specific rubric to assess the performance. Once this is decided, a decision to use an analytic or holistic rubric should be made <sup>[10]</sup>. After these decisions, the process of defining attributes on which to grade the performance or product begins <sup>[11]</sup>. Detailed steps for considering the learning objectives, defining attributes and attribute characteristics, and distinguishing levels of work for each attribute are defined in Figure 1 below.

- Step 1: Re-examine the learning objectives to be addressed by the task.
- Step 2: Identify specific observable attributes that you want to see (as well as those you don't want to see) your students demonstrate in their product, process, or performance.
- Step 3: Brainstorm characteristics that describe each attribute.

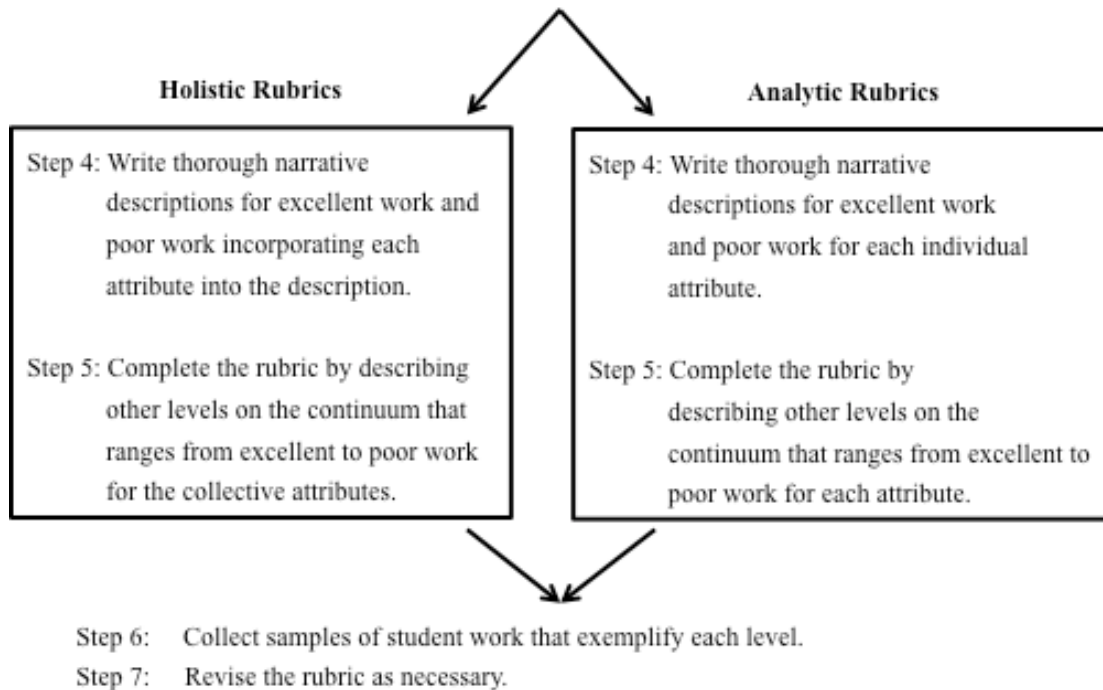


Figure 1. *Rubric Development Steps in Academia in Designing scoring rubrics for your classroom.* (p. 6) by Craig A. Mertler. *Practical Assessment, Research & Evaluation*, 7(25). 2001.

Once the rubric is developed, issues of validity and reliability must be addressed.

### *Scoring Rubrics: Validity and Reliability*

Validity and reliability are key issues in determining whether a rubric is appropriate for use in evaluation. Reliability, a prerequisite for validity, refers to the consistency of assessment scores; validity refers to the degree to which interpretations of scoring are correct and appropriate<sup>[12]</sup>.

Moskal and Leydens<sup>[12]</sup> describe three types of evidence to support the validity of a rubric: content, construct, and criterion-related evidence. Content-related evidence refers to how much a student's assessment response reflects the student's knowledge of the content area. Construct-related evidence refers to a student's reasoning process for performing a task or solving a problem. Criterion-related evidence refers to the extent the results of an assessment correlate with current or future events. Criterion-related evidence are commonly found in engineering courses where classes mimic the work environment of practicing engineers<sup>[12]</sup>.

Validity criterion are generally decided at the beginning stages of rubric development, when the assessment objectives are determined. Once the objectives are determined, the type of evidence used for evaluation can then be taken into consideration.

Once rubrics are developed and in use, issues of reliability are of concern. To ensure scoring is consistent across students, interrater reliability is often used to determine to what extent evaluators align in their scores. Clear scoring rubrics can minimize scoring differences because they formalize the criteria at each scoring level <sup>[12]</sup>.

## **Methods: Translating Rubric Development to Industry**

The process for developing rubrics in industry closely parallels the method for rubric development in academia. This section describes the need for rubrics in industry, and explores a method for developing a rubric to assess the professional skills of engineers. The rubric is designed to evaluate professional skills in the context of playing team-building games during an engineering hiring process. First is a brief background on the need for rubrics in industry, followed by a proposed process for developing and using a rubric for assessing professional skills in this context.

### *Understanding the Need for Rubrics in Industry*

Growth in engineers' careers relies on promotions and hiring based on evaluations at all levels. Many companies have clearly defined values that guide company employees on behaviors and qualities they should exhibit. These values guide the hiring and promotion processes, as evaluators look to see which employees exhibit desired traits; however, many companies do not elaborate on the definitions to further clarify quality scales on the different traits. This can cause issues in reliability of evaluations, which may hinder an engineer's chance for employment or promotion.

A recent trend among engineering companies has been to incorporate team-building games to evaluate professional skills, in addition to behavioral and technical interviews. Potential employees in on-site interviews are asked to play a series of games that may require teamwork and/or negotiation in one or several units. The goal is for evaluators to see how candidates behave and exhibit professional skills in a spontaneous and more 'natural' environment, where canned responses cannot be given <sup>[2]</sup>. To judge candidates, several evaluators are assigned participants to watch and are given definitions of traits to look for (e.g., 'Communication – the participant communicates clearly to his or her team') and given a scale of 1 (poor) to 5 (excellent) to rank participants. At the end of the games, the scores of each participant are averaged for a final score towards their hiring package. This process is subject to a two key issues: First, because the definitions are not based in the context of the game, it can be difficult for evaluators to judge if an action counts as a definition or not. For the communication example, if a participant writes down an idea clearly, but struggles to articulate it, what score should be given? Second, because there is no sliding scale for quality, evaluators must base their scores on their subjective understandings of a definition. If two evaluators actually agree upon the actions a participant did but end up giving the participant differing scores of 3 and 4, how do they account for discrepancies in scoring?

The need for detailed rubrics also extends past the hiring process to the promotion of engineers. A lack of clearly defined behavior traits can make it difficult for supervisors to acknowledge progress of an employee, and for employees to keep track of their growth in a company. For example, Figure 2 below displays NASA’s leadership model for managers <sup>[13]</sup>. The official website and PDF handout of this leadership model provides detailed definitions of each desired trait (each of which is commendable) but fails to define quality levels of performance for the best of worst examples of demonstrating a trait. This allows for vague feedback to be given during performance reviews, and can be cause for misunderstandings about behavior between supervisors and employees. Behavior that is viewed as ‘assertive’ and ‘leader-like’ by men is often construed as ‘aggressive’ and ‘bossy’ when exhibited by women <sup>[14]</sup>; detailed rubrics may assist in clearing up gender and race biases by allowing employees to explain behaviors in the context of the rubric and clarify potentially negative stereotypes. When rubrics are constructed and evaluated by employees in collaboration with managers, they can also be used as a learning tool, as employees must reflect on their professional skill development.



Figure 2. NASA Leadership Model for Managers. [13]

### *Developing Rubrics in Industry*

As with rubric development in academia, the first step in rubric development in industry is to determine the type of rubric to be developed. Context is critical when determining which



qualities are best to evaluate in particular games. Certain games selected may be more appropriate for evaluating for particular traits; for example, a game requiring a group as a whole to accomplish a task may be best for evaluating teamwork and communication skills, while a game requiring negotiation among group members might be best for evaluating flexibility. For this reason, the rubrics developed must be task-specific (doing a particular action in the context of the game) and analytic (each criterion is evaluated on its own). Figure 2 below demonstrates a modified rubric development process that mirrors the one used in academic settings.

The following set of steps details the process for rubric development industry. Examples are provided for a game called ‘Card Pieces,’ and a sample portion of a rubric is presented after the description of the process in Table 1.

### *Sample Game: Card Pieces*

*Game Description:* Participants are split into groups of 2-3 people. Each group is given an envelope that contains cut up pieces of 4 playing cards such that no single envelope contains the parts to make one whole card. The objective of the game is for teams to barter with each other in order to become the first team to piece together the 4 cards. First, each team is given 3 minutes to sort its pieces, determine which pieces it needs to make complete cards, and develop a bargaining plan. Teams are then allowed to start bartering for pieces. People can barter on their own or as a team. When the time is up, whichever team has the most number of completed cards wins.

Step 1: Identify observable attributes to evaluate. To determine the specific criterion to evaluate in each game, interviews may be conducted with hiring experts who have experience using team-building games for hiring. Interviewees should be given descriptions of the games that will be used and asked to identify the top two or three professional skills they believe would be best to evaluate in this game. For the game ‘Card Pieces,’ an example attribute experts may want to observe is flexibility of the participants.

Step 2: List characteristics that describe each attribute. Once the list of attributes is compiled, each attribute should be described or defined in the context of the game. For ‘Card Pieces,’ flexibility may be defined as the ability to compromise in exchanging cards with other teams, or the ability to change strategies when discussing amongst teammates.

Step 3: Write thorough descriptions of behaviors that constitute the best and worst demonstration of each attribute. Examples should be as specific as possible in order to be easily identifiable during game play. An example of a best behavior for flexibility during ‘Card Pieces’ may be “the participant willingly engages in exchanging his or her own card pieces without causing any negotiations to break down or agreements to be broken.” An example of a worst behavior for flexibility may be “the participant refuses to compromise or discuss an exchange of cards.”

Step 4: Complete the rubric by describing other levels on the continuum that ranges from best to worst demonstration for each attribute. Since engineering candidates are evaluated while playing the games in real-time, industry experts advise not making the rubric too complicated [2]. For this reason, only 2 to 3 quality scale descriptions should be used in total. For ‘Card Pieces,’ an

example of a good demonstration of flexibility may be “a participant engages in card exchange but is not the one to initiate negotiations at any point in the game.”

Step 5: Revise the rubric as necessary. Engage with all potential evaluators to review the attributes and descriptions and clarify any confusing language.

PROFESSIONAL SKILL	RATING		
	POOR (1 pt.)	GOOD (2 pts.)	EXCELLENT (3 pts.)
<b>Flexibility</b> - the ability to compromise in exchanging cards with other teams	A participant refuses to compromise or discuss an exchange of cards.	A participant engages in card exchange but is not the one to initiate negotiations at any point in the game.	A participant willingly engages in exchanging his or her own card pieces without causing negotiations to break down or agreements to be broken.

Table 1. A sample rubric demonstrating one attribute and characteristics describing the quality scale from poor to excellent.

Once the rubrics are completed for all games, a practice round of the games should be played and evaluators should practice evaluating with the rubric before implementing the rubric during the hiring process. Interrater reliability should be conducted to ensure all evaluators are consistent in their scoring; any inconsistencies should be addressed and modifications to the rubric should be made. This practice process should be repeated until an acceptable interrater reliability score (based on the number of evaluators and percentage agreement statistically calculated) is reached.

### *Ongoing and Future Research*

The rubric development process described above is currently being validated and tested with three hiring games used by a company in the manufacturing and automotive engineering industry. Future research and publications will detail the implementation of the rubric development process in an industrial setting, including (1) results of evaluating engineers’ professional skills with the new rubrics, (2) comparison of these scores with original rubric scores and usability feedback from evaluators, and (3) cost-benefit analyses of implementing these rubrics industry.

### **Conclusion**

This paper has articulated a process for developing a behavior-based rubric for use in engineering industry hiring processes involving team-building games. The development process can be extended past hiring in industry to the use of evaluations for the promotions of engineers. Implications for developing detailed behavior-based rubrics in industry are three-fold: first, bias

in evaluations for hiring can be reduced by providing clear, contextual definitions of professional skills in games; second, rubrics can be used as a learning tool when constructed and evaluated by employees in collaboration with supervisors; and third, detailed rubrics can help explain cultural, gender, and racial behaviors that may be perceived negatively in the context of the rubric.

## References

- [1] National Academy of Engineering, "The engineer of 2020: Visions of engineering in the new century," *Natl. Acad. Eng. Wash. DC*, 2004.
- [2] Engineering Human Resource Manager, "Personal Communication," 2014.
- [3] R. Shelby, E. Patten, F. Ansari, L. Pruitt, G. Walker, and J. Wang, "Implementation of Leadership and Service Learning in a First-Year Engineering Course Enhances Professional Skills," *Int. J. Eng. Educ.*, vol. 29, no. 1, pp. 1–14, 2013.
- [4] L. J. Shuman, M. Besterfield-Sacre, and J. McGourty, "The ABET 'Professional Skills' - Can They Be Taught? Can They Be Assessed?," *J. Eng. Educ.*, vol. 94, no. 1, pp. 41–55, Jan. 2005.
- [5] J. Prus and R. Johnson, "A critical review of student assessment options," *New Dir. Community Coll.*, vol. 1994, no. 88, pp. 69–83, 1994.
- [6] H. G. Andrade, "Teaching With Rubrics: The Good, the Bad, and the Ugly," *Coll. Teach.*, vol. 53, no. 1, pp. 27–31, Jan. 2005.
- [7] Y. M. Reddy and H. Andrade, "A review of rubric use in higher education," *Assess. Eval. High. Educ.*, vol. 35, no. 4, pp. 435–448, 2010.
- [8] R. Stiggins, J. A. Arter, J. Chappuis, and S. Chappuis, "Classroom assessment for learning," *Doing It Right-Using It Well*, 2007.
- [9] S. M. Brookhart, *How to create and use rubrics for formative assessment and grading*. Ascd, 2013.
- [10] S. Bargainnier, "Fundamentals of rubrics," *Pac. Crest*, pp. 1–4, 2003.
- [11] C. A. Mertler, "Designing scoring rubrics for your classroom," *Pract. Assess. Res. Eval.*, vol. 7, no. 25, pp. 1–10, 2001.
- [12] B. M. Moskal and J. A. Leydens, "Scoring rubric development: Validity and reliability," *Pract. Assess. Res. Eval.*, vol. 7, no. 10, pp. 71–81, 2000.
- [13] NASA, .
- [14] H. Ibarra, R. Ely, and D. Kolb, "Women rising: The unseen barriers," *Harv. Bus. Rev.*, vol. 91, no. 9, pp. 60–66, 2013.