ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.245

Measurement and improvement of the reliability in the competency assessment of the engineer's training.

Juan Victor Bernal - Olvera, Ph. D[1], Mónica Belem Bernal - Perez, MBA[2],and Mireya Berenice Monroy - Anieva, MBA[3]
[1,2,3]*Tecnológico Nacional de México/TES Cuautitlán Izcalli, Mexico, juan.bo@cuautitlan.tecnm.mx, monica.bp@cuautitlan.tecnm.mx, mireya.ma@ cuautitlan.tecnm.mx,*

Abstract– In this paper, the importance of a properly designed rubric in the competency-based assessment process is emphasized. The consistency of this rubric is crucial to ensure a fair evaluation by the facilitator. The research focused on developing and applying a rubric to evaluate competencies in the training of engineers within a university in Mexico. The study used Cronbach's alpha and the Pearson coefficient to quantify the correlation between the criteria and parameters. Particularly, the results in this rubric showed low values for the reference indicators, displaying a lack of consistency. By reviewing the correlation between the criteria and parameters, alternatives are proposed for improvement until reaching the appropriate value that provides security and certainty in its application.

Keywords-- Summative evaluation, engineering education, academic standards, proficiency-based education, student evaluation.

I. Introduction

The evaluation system in the training of engineers is of utmost importance and must be reliable to ensure a fair and accurate assessment of the competence level of the students. Building instruments for measuring the degree of acquisition of knowledge, skills, and abilities can be challenging, but it is possible to create reliable tools that meet the standards required for quality processes. It is crucial to have a consistent rubric in place that can accurately evaluate the competencies and provide a fair assessment.

Thus, the planning phase in ISO 9001:2015 [1] is crucial as it emphasizes the need to determine the risks and opportunities of the institution and its interested parties, which is essential to ensure a controlled and monitored production and provision of service. The IATF 16949:2016 standard also highlights the importance of analyzing measurement systems within Chapter 7 [2]. The Six Sigma quality improvement tool, with its DMAIC methodology, is a widely used approach to improve existing processes and products. It is formed with the onomatopoeia of the first letter of the Define, Measure, Analyze, Implement and Control phases. The tool places great importance on measurement systems during its second stage to ensure that the values that arise from the process are reliable and accurate.

A piece of knowledge can be measured through different instruments, but the criteria used must have the characteristic of being transparent [3]. In this sense, transparency means that educators are explicit about their assessment expectations and students know what they need to demonstrate about their learning [4]. Failure to comply with these characteristics can have negative consequences for the generation of knowledge in students, deficiencies in the teacher-student relationship, and non-representative assessments of the level of acquisition of competence.

It is important to have a clear and reliable measurement system for assessing the knowledge of engineers during their education and training. Any flaws in the system can lead to two extreme scenarios. Firstly, it can result in a significant drop in student enrollment due to academic issues of unapproved subjects, which can put the existence of the institution at risk. Secondly, it can lead to students with a low level of knowledge, skills, and abilities, which may not meet the requirements demanded by the area in which they develop professionally. Both scenarios are equally hazardous as they can damage the reputation of the University and, in the medium or long term, threaten its existence as a teaching organization.

A. The evaluation system

A well-designed evaluation method is crucial to ensure a fair and accurate assessment of competencies. Formats with analytical criteria have been proven to be more effective than conventional lists of requirements, as highlighted by Nadolski, Hummel, Rusman, et al [5]. Tejada and Ruiz [6] propose a process that sizes the evaluation by competencies, which considers what, how, when, for what, who, and with what to evaluate. The analysis focuses on the instruments, techniques, or devices for collecting information, which are necessary to generate a result of the degree of acquisition of a competence. Choosing the right evaluation tools is essential to ensure that the results are reliable and accurate.

The assessment instrument or rubric is a crucial element in the evaluation process. It is a coherent set of criteria that guides the review of student work and includes a description of levels of quality of performance, as highlighted by Brookhart [7]. Andrade [8] and, Arter and Chappuis [9] point out that rubrics allow for the articulation of expectations for students' work by listing rules and descriptions of the level of performance throughout a continuous quality process. Moreover, rubrics are suitable for use in both formative and summative evaluations as they contain descriptions of compliance with requirements and do not assign only an evaluation grade, as emphasized by Brookhart and Chen [10]. Rubrics provide a clear and objective way to evaluate performance and provide constructive feedback to students to improve their work.

Rubrics are multifaceted instruments that can be used to improve learning and assessment practices for various competencies. Teachers often reuse existing rubrics, but they also create new ones to suit their specific needs, as highlighted by Rusman and Dirkx [11]. However, there are some potential errors that can occur when applying rubrics as structured evaluation guides, as pointed out by Szfaran [12]. The first error can happen when a different rater is hired to judge a discrepancy in scores reported by an initial rater. This can introduce inconsistency and bias in the evaluation process. The second error is discontinuing training before raters have demonstrated a satisfactory level of interrater reliability. This can lead to inconsistent application of the rubric and unreliable results. It is essential to ensure adequate training and interrater reliability to maintain the consistency and accuracy of the evaluation process.

The rubric is a unique assessment tool that considers descriptions of criteria and levels of performance to provide a reliable and objective evaluation of student work, as highlighted by Brookhart [13]. However, there are some potential difficulties in expressing certain aspects of the rubric reliably during its use, as pointed out by Bearman and Ajjawi [4]. Firstly, there is a type of knowledge, such as soft skills, that cannot be expressed accurately using rubrics. Secondly, the criteria used in the rubric may not be clear to the person carrying out the evaluation, leading to inconsistencies in the evaluation process. Lastly, making some criteria transparent can cause other criteria to become opaque and less transparent. It is crucial to be aware of these potential issues and take steps to minimize their impact on the evaluation process.

B. *Consistency requirements in the measurement system*

Brookhart and Chen [14] highlight the importance of different reliability indicators that measure consistency between several factors, such as interpretation and use by evaluators, criteria that have different definitions of consistency, absolute or relative, and the use of different calculation methods. Consistency between evaluators is essential to ensure the quality of students' performance, as emphasized by Bernal, Bernal, and Monroy [15]. The measures agreed upon must be right for the common educational model of teaching according to standards, objectives, or goals of the course and the training program. It is essential to ensure that the evaluation tools and methods used are dependable and valid for the specific context and competencies being assessed. Adequate training and interrater reliability can help to ensure consistency between evaluators and improve the accuracy and reliability of the evaluation process.

Internal consistency is a critical characteristic that can be evaluated to decide how well the elements that make up surveys or tests designed to measure the same construct actually do so, as highlighted by Minitab [16]. A construct is a variable within a hypothesis, theory, or theoretical model, such as a theme, characteristic, or underlying skill, which is related to other attributes like reading comprehension or customer satisfaction, as explained by Hernández and Mendoza [17]. Cohen's kappa and inter-class correlation are some indicators that are used to measure consistency and make judgments about one or more factors at random. On the other hand, Cronbach's alpha is an indicator that measures the consistency of the set of factors through the correlation that exists between them as a whole. It is commonly used to validate perception surveys in business or some educational fields. It is essential to choose the right indicator to measure consistency depending on the context and purpose of the evaluation.

C. *Cronbach's alpha*

Cronbach's alpha is denoted by the Greek letter α and is calculated using a multiplicative function between the ratio of the degrees of freedom and a unit rate decreased by a relationship of variances between the criteria that make up the questionnaire and the grades assigned by the evaluators, as shown in Equation 1.

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum V_{subtests}}{V_{test}}\right] \quad (1)$$

Where n is the number of subtests or criteria, $\sum Vsubtests$ is the sum of the variance of all the subtests, and $Vtest$ is the variance of all the tests evaluated. Cronbach's alpha is a widely used indicator of internal consistency, and it measures how well the items in a questionnaire or test measure the same construct. A high value of Cronbach's alpha indicates high internal consistency among the items in the questionnaire or test, while a low value shows poor internal consistency, which means that the items are not measuring the same construct. It is essential to calculate Cronbach's alpha for any questionnaire or test to ensure that the items have high internal consistency and accurately measure the construct being evaluated.

For interpretation purposes, Norcini suggests that a value of 0.80 is acceptable for Cronbach's alpha [18]. On the other hand, Iacobucci and Duhachek suggest that values of α equal to 0.70 can be accepted as a first level in research [19]. Cronbach sets up that the use of Cronbach's alpha considers the dispersion of the data [20]. Cronbach's alpha is the mean of all possible coefficients divided by halves and is considered as a lower limit for the coefficient of precision of the estimates and for the proportion of the test variance attributable to common factors between the items. It is essential to keep in mind that the interpretation of Cronbach's alpha values depends on the context and purpose of the evaluation. In general, higher values of alpha indicate higher internal consistency among the items in the questionnaire or test.

Cronbach's alpha has three control variables: the number of subtests (n), the sum of the variances of the set of subtests or criteria that make up the evaluation instrument ($\sum Vsubtests$), and the variance of all the tests applied to the participants ($Vtest$). By performing a limits analysis, it is possible to determine which variable should be changed to achieve better results in the value of alpha. If n tends to 0, the limit is zero, which means that without any subtest numbers, there is no data. On the other hand, if n tends to infinity, the limit of Equation 1 is $\left[1 - \frac{\sum V_{subtests}}{V_{test}}\right]$.

If $\sum Vsubtests$ is reduced to zero, then α takes the value of $\frac{n}{n-1}$, which forms a graph with a vertical asymptote at n=1. But if it increases to infinity, α also tends to minus infinity (-∞). Finally, if $Vtest$ tends to zero, α tends to at least infinity. However, if it increases to infinity, α is equal to $\frac{n}{n-1}$. It is essential to consider these variables when applying Cronbach's alpha to ensure accurate and reliable results. By analyzing these variables, it is possible to determine which variable should be adjusted to improve the value of alpha and increase the internal consistency of the questionnaire or test.

D. *Research context*

This study was conducted within the Tecnológico Nacional de México, on its campus of the Tecnológico de Estudios Superiores de Cuautitlán Izcalli (TESCI), which is located 30 km northeast of Mexico City. This school is a decentralized public institution of the Government of the State of Mexico that offers 8 engineering degrees, a degree in Public Accounting, and 2 postgraduate programs with a focus on technological research. It is worth noting that 7 of the 8 engineering programs have international accreditations, which is a testament to the high quality of education provided by the institution. The campus was inaugurated in 1998 and currently has an enrollment of more than 5,000 students. The infrastructure of the campus includes classroom buildings, laboratories, green patios, sports, and recreation areas, which provide a conducive learning environment for the students.

The Industrial Engineering degree is the division with the largest number of students, comprising around 20% of the total student population. The students in this program are divided into two shifts, morning, and afternoon. The teaching-learning processes in this program have a competency-based training approach, which is a modern and effective method of education. Moreover, the teaching staff of the program is receiving constant training in this andragogical training to enhance their teaching skills and ensure that they are up to date with the latest trends and practices in the field of education. The courses taught in this program have a duration of six months, and the academic year is divided into two periods. The first period takes place from February to July, and the second period takes place from August to January of the following year.

II. METHODOLOGY

The research involves the construction of an Activity Report rubric and an experimental development. The scientific method is used in the study, and the following steps are proposed:

1. Problem Formulation: The first step involves identifying and defining the problem, which is related to the evaluation process of the Activity Report rubric.
2. Preparation of the Testing Hypothesis: Based on the problem formulation, hypotheses are formulated to improve the evaluation process.
3. Design and Execution of the Validation Experiment: The hypotheses are tested through experimentation, and data is collected to evaluate the effectiveness of the proposed changes.
4. Obtaining and Analyzing Results and Discussion: The data collected is analyzed to draw conclusions about the effectiveness of the proposed changes. The results and analysis are then discussed to provide a comprehensive understanding of the research findings.
5. Proposed Solution: Based on the analysis and discussion, a proposed solution is presented to address the identified problem and improve the evaluation process.
6. Conclusions and Recommendations: The research concludes with a summary of the findings and recommendations for future research.

These steps provide a structured and systematic approach to conducting research, which ensures that the results obtained are reliable and valid. By following these steps, researchers can develop a comprehensive understanding of the problem and propose effective solutions to address it.

The construction of the Activity Report rubric is an essential step in the research process, as it provides a clear and objective evaluation tool. The experimental development allows for the testing of hypotheses and the evaluation of the proposed changes. Overall, the scientific method provides a structured and systematic approach to conducting research, which ensures that the results obtained are reliable and valid.

A. *Methodology for developing a rubric*

A methodology is established to build a rubric, based on the work developed by Del Pozo [21] and Tejada and Ruíz [6]. The methodology consists of six stages, which are shown in Figure 1 and described below:

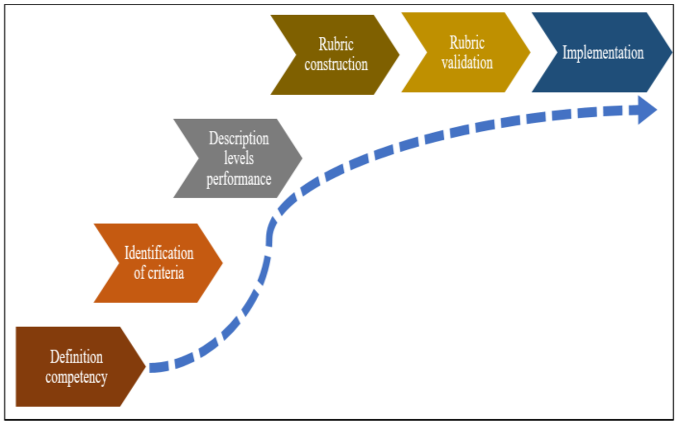

Fig. 1 Methodology for constructing an analytical rubric.

1) Definition of the Competency: The first stage involves defining the competency that will be evaluated through the rubric. This stage requires a clear understanding of the competency and its related knowledge, skills, and abilities.

2) Identification of Criteria: The second stage involves identifying the criteria that will be used to evaluate the competency. The criteria should be specific and measurable to ensure a clear evaluation.

3) Description of Levels of Performance: The third stage involves describing the different levels of performance for each criterion. This stage requires a clear understanding of the criteria and their respective levels of performance.

4) Construction of the Rubric: The fourth stage involves constructing the rubric based on the identified criteria and levels of performance. The rubric should be clear, concise, and objective to ensure a fair evaluation.

5) Validation of the Rubric: The fifth stage involves validating the rubric by testing it on a sample of participants. The feedback obtained from the participants can be used to refine the rubric and improve its effectiveness.

6) Implementation of the rubric: The final stage involves implementing the rubric in the evaluation process. The rubric should be used consistently and objectively to ensure a fair evaluation.

B. *Problem Statement*

It is essential to ensure that evaluations are fair and consistent in determining the acquisition of competencies in engineering training programs. In this regard, it is important to verify or establish the consistency of the rubric, which serves as a structured evaluation guide. Typically, teachers who belong to different departments propose rubrics that are used throughout the semester to assign grades to the products submitted during competency-based learning. However, even when there is agreement among the Academy members who use it, the question still remains whether the rubric is consistent enough to measure the level of competency acquisition accurately. Therefore, it is crucial to ensure that the rubric used is consistent and reliable to support fair evaluations.

C. *The test rubric*

It is important to keep in mind that assessment should be incorporated into the planning process, taking into consideration the expected learning outcomes [22]. One approach to establishing learning competencies is through the creation of an Activity Report, which can be utilized in various subject areas. The goal of this report is to encourage students to analyze a specific topic and produce written evidence that demonstrates their ability to solve a problem, conduct a critical analysis, or synthesize related terms in a coherent manner that is relevant to the topic at hand.

To assess professional learning activities, 8 criteria are recommended, including cover or delivery format, spelling, delivery on time, use of bibliography, objectives, introduction, development, and conclusion. Each criterion is assessed using the Likert scale, with 3 parameters assigned to determine the merit of each criterion. The most comprehensive criterion is assigned a value of 3, while partial compliance results in a score of 2. If a criterion is not met, or the minimum requirements are not met, a score of 1 is assigned. For example, the first criterion evaluates compliance with the cover page or delivery format, which are standard requirements for report submissions. If all the general data requested is included, a score of 3 is assigned. However, if 1 or 2 items are omitted, a score of 2 is given, and if more than 3 items are missing, a score of 1 is assigned. Table 1 provides a detailed description of each criterion and its parameters for assessment.

TABLE I
CURRENT RUBRIC

Criteria	Scale		
	3	2	1
1. Coverand delivery format.	The student uses the format and writes down all the general data	The student uses the format but omits 1 to 2 general facts.	The student does not use the format or omits more than 3 general data.
2. Orthography.	The report has not ortography mistakes.	The report has 1- 5 ortography mistakes.	The report has more than 5 ortography mistakes.
3. Delivery on time.	The report was delivered within the established deadline.	-	The report was delivered after the established deadline.
4. Use of bibliography.	The report displays bibliography references using APA format.	-	The report does not displays bibliography references.
5. Objective.	The report has an objective, and this is related to the proposed application.	-	The report does not have an objective, or this is not related to the proposed application.
6. Introduction.	The information presented is relevant and introduces the topic being applied at 100%in at least a 5-line paragraph.	The information presented is relevant and introduces the topic being applied in more than /0% but less than 100% in at least a 5-line paragraph.	The information presented is relevant and introduces the topic being applied at less than 70%
7. Development.	The report contains detailed and orderly reasoning, using the appropriate strategy following the steps to apply the knowledge 100%	The report contains detailed and orderly reasoning, using the appropriate strategy following the steps to apply the knowledge correctly at	The report contains detailed and orderly reasoning, using the appropriate strategy following the steps to apply the knowledge correctly at
8. Conclusion	The report presents a conclusion verifying the application of knowledge, recording differences or recommendations in at least one 5-line paragraph.	The report presents a conclusion but it was not verifying the application of knowledge, or recording differences and recommendations in at least one 5-line paragraph.	The report does not present conclusions or these lack value regarding the application

D. *Testing hypothesis*

It's important to keep in mind that the reliability of an evaluation instrument is determined by its Cronbach's alpha value, which should ideally be at least 0.7. In case the value is lower than this threshold, it may be necessary to make some adjustments to ensure that the instrument meets the reliability requirements.

E. *Design and execution of the validation experiment*

In the first semester of 2022, a rubric was utilized to evaluate ten randomly selected products submitted by students in the Economics subject of the Industrial Engineering program. The evaluations were conducted through the Microsoft Teams platform. The professor responsible for the evaluations possesses the necessary qualifications and experience to conduct the experiment effectively. The teacher has previously taught the subject, holds a master's degree, and has been teaching competency-based engineering training programs for over five years.

The designated teacher reviews the reports and assigns a weight of 3, 2, or 1 depending on the degree of content compliance with the rubric requirement. The results of the evaluations are then arranged in a matrix format, where the tests, denoted by $test_i, where\ i = 1, 2, \ldots, 10$, form the rows, and the rubric criteria, denoted by $subtest_i,\ where\ i = 1, 2, \ldots, 8$, are listed in Table 2.

TABLE 2
RESULTS OBTAINED WHEN APPLYING THE CURRENT RUBRIC

Applied to	Subtest 1	Subtest 2	Subtest 3	Subtest 4	Subtest 5	Subtest 6	Subtest 7	Subtest 8
Test 1	3	1	3	3	3	3	3	2
Test 2	3	3	1	3	3	2	3	3
Test 3	3	1	3	3	3	3	2	3
Test 4	3	3	1	3	3	2	3	3
Test 5	2	3	3	3	3	2	3	2
Test 6	3	3	3	3	3	2	3	3
Test 7	3	1	3	3	3	3	3	3
Test 8	3	2	1	3	3	2	2	3
Test 9	3	2	1	3	3	2	2	3
Test 10	1	2	1	3	3	1	1	1

III. RESULTS AND DISCUSSION

To obtain the Cronbach's alpha, calculations were performed using Equation 1 with the support of an Excel sheet. The results were then verified using Minitab software, which yielded a value of 0.6165, with a variation of less than 5% compared to Equation 1. These results are presented in Table 3. It can be observed that the calculated value of α is less than 0.7, indicating that the consistency of the rubric is insufficient to apply it in competency evaluations within engineering training programs.

TABLE 3
RESULTS OF THE CONSISTENCY OF THE CURRENT RUBRIC USING CRONBACH'S ALPHA

Applied to	Subtest 1	Subtest 2	Subtest 3	Subtest 4	Subtest 5	Subtest 6	Subtest 7	Subtest 8	SUM
Test 1	3	1	3	3	3	3	3	2	21
Test 2	3	3	1	3	3	2	3	3	21
Test 3	3	1	3	3	3	3	2	3	21
Test 4	3	3	1	3	3	2	3	3	21
Test 5	2	3	3	3	3	2	3	2	21
Test 6	3	3	3	3	3	2	3	3	23
Test 7	3	1	3	3	3	3	3	3	22
Test 8	3	2	1	3	3	2	2	3	19
Test 9	3	2	1	3	3	2	2	3	19
Test 10	1	2	1	3	3	1	1	1	13
Variance	0.455556	0.76667	1.11111	0	0	0.4	0.5	0.48889	

Sum of variances	3.7222
Variance of the sum of the tests	7.6556
n=	8
α =	0.6

A. *Discussion*

The lack of consistency in the rubric is reflected in the Cronbach's alpha values, which are less than 0.7 for most of the criteria or subtests, except for the number two. The statistics for subtest 4 and subtest 5 were omitted from the analysis as they showed a variation of zero. These results are presented in Table 4.

TABLE 4
PEARSON CORRELATION AND CRONBACH'S ALPHA WITH THE ORIGINAL CRITERIA OF THE RUBRIC

Omitted variable	Adjusted total mean	Adjusted total sta dev	Total correlation adjusted by element	Square Multiple Correlation	Cronbach's alpha
Subtest 1	11.400	2.271	0.6670	0.9268	0.4580
Subtest 2	12.000	2.828	-0.2243	0.9307	0.7882
Subtest 3	12.100	2.331	0.2261	0.7333	0.6493
Subtest 6	11.900	2.378	0.5319	0.9621	0.5157
Subtest 7	11.600	2.171	0.7963	0.8846	0.3950
Subtest 8	11.500	2.321	0.5476	0.8864	0.5000

To improve the consistency, it is suggested to review the Pearson correlations between the $Subtest_i$ elements as shown in Table 5. It can be observed that there is no correlation defined in the columns of Subtest 4 and Subtest 5, which makes it difficult to establish any variation since all have been rated 3 as presented in Table 2. Therefore, it is recommended to add an intermediate parameter in criteria 4 and 5. Similarly, criterion 3 needs to be revised as it has a low correlation value of 0 with subtest 8. This will help in improving the correlation with criteria 6 and 7.

TABLE 5

PEARSON CORRELATION BETWEEN CRITERIA OF THE CURRENT RUBRIC

	Subtest 1	Subtest 2	Subtest 3	Subtest 4	Subtest 5	Subtest 6	Subtest 7
Subtest 2	-0.132						
Subtest 3	0.156	-0.361					
Subtest 4	*	*	*				
Subtest 5	*	*	*	*			
Subtest 6	0.677	-0.642	0.667	*	*		
Subtest 7	0.582	0.269	0.447	*	*	0.497	
Subtest 8	0.895	0.073	0.000	*	*	0.452	0.449

B. *Improvement and application proposal*

The changes in the rubric for criteria 3, 4, and 5 are quite significant. For instance, criterion 3 now includes a parameter 2, which allows for a 12-hour extension beyond the deadline for the activity to be submitted. However, if the submission is made after this extended time, it will be rated as 1. Similarly, criterion 4 now requires references to be shown, but it is not necessary to follow the APA format. Not including any references will lead to a rating of 1. Moreover, criterion 5 now involves assigning a rating of 2 if the objective is not relevant to the topic, and a rating of 1 if it is not included at all. Please refer to Figure 2 for a better understanding of these changes.

Criteria	Scale		
	3	2	1
3. Delivery on time.	The report was delivered within the established deadline.	The report was delivered after the established deadline but before 12-hours late.	The report was delivered more than 12 hours after the established deadline.
4. Use of bibliography.	The report displays bibliography references using APA format.	The report displays bibliography references without using APA format.	The report does not displays bibliography references.
5. Objective.	The report has an objective, and this is related to the proposed application.	The report does not have an objective, or this is not related to the proposed application.	The report does not have an objective.

Fig. 2 Additions to the current rubric.

It seems that after reviewing the same reports again, with an emphasis on criteria 3, 4, and 5, and applying parameters 1 to 3, the new results have been presented in Table 6. The results are highlighted in bold lettering and indicate an improvement in Cronbach's alpha, with a value of 0.8016 supported by Minitab.

TABLE 6

RESULTS OF CRONBACH'S ALPHA WITH THE MODIFIED CRITERIA IN THE RUBRIC

Applied to	Subtest 1	Subtest 2	**Subtest 3**	**Subtest 4**	**Subtest 5**	Subtest 6	Subtest 7	Subtest 8	SUMA
Test 1	3	1	**3**	**3**	**2**	3	3	2	20
Test 2	3	3	**2**	**2**	**3**	2	3	3	21
Test 3	3	1	**3**	**2**	**3**	3	2	3	20
Test 4	3	3	**2**	**3**	**2**	2	3	3	21
Test 5	2	3	**3**	**2**	**2**	2	3	2	19
Test 6	3	3	**3**	**3**	**3**	2	3	3	23
Test 7	3	1	**3**	**3**	**3**	3	3	3	22
Test 8	3	2	**2**	**2**	**2**	2	2	3	18
Test 9	3	2	**1**	**2**	**2**	2	2	3	17
Test 10	1	2	**1**	**2**	**1**	1	1	1	10
Variance	0.4556	0.7667	0.6778	0.2667	0.4556	0.4	0.5	0.4889	

Sum of variances	4.0111
Variance of the sum of the tests	13.433
n=	8
α =	0.8016

The correlation statistics show that all the criteria are now present in Table 7, and the consistencies have improved, meeting the standards set by Norcini [18] and Iacobucci and Duhacheck [19] for Cronbach's alpha.

TABLE 7

STATISTICS OF THE MODIFIED RUBRIC CRITERIA AND CRONBACH'S ALPHA

Omitted variable	Adjusted total mean	Adjusted total sta dev	Total correlation adjusted by element	Square Multiple Correlation	Cronbach's alpha
Subtest 1	16.400	3.134	0.7459	0.9337	0.7443
Subtest 2	17.000	3.682	-0.1379	0.9632	0.8874
Subtest 3	16.800	3.120	0.5883	0.8645	0.7671
Subtest 4	16.700	3.401	0.4555	0.6730	0.7890
Subtest 5	16.800	3.120	0.7704	0.8086	0.7405
Subtest 6	16.900	3.281	0.5461	0.9790	0.7754
Subtest 7	16.600	3.062	0.8210	0.9396	0.7299
Subtest 8	16.500	3.171	0.6515	0.9306	0.7580

The Pearson correlation matrix presented in Table 8 indicates that all correlations are positive, except for criterion or Subtest 2. It seems that this exception can be attributed to the fact that there is a low possibility of finding spelling errors with on-time deliveries, the use of bibliography, objectives, and introduction.

TABLE 8

PEARSON CORRELATION BETWEEN CRITERIA IN THE MODIFIED RUBRIC

	Subtest 1	Subtest 2	Subtest 3	Subtest 4	Subtest 5	Subtest 6	Subtest 7
Subtest 2	-0.132						
Subtest 3	0.380	-0.200					
Subtest 4	0.383	-0.098	0.470				
Subtest 5	0.707	-0.056	0.620	0.255			
Subtest 6	0.677	-0.642	0.726	0.408	0.625		
Subtest 7	0.582	0.269	0.668	0.609	0.582	0.497	
Subtest 8	0.895	0.073	0.232	0.185	0.753	0.452	0.449

IV. Conclusions and Recommendations

It's important to keep in mind that a rubric is a crucial tool for evaluating skill acquisition in engineering training programs. To ensure its proper use, there are two key conditions that must be met: first, it should be reviewed by an expert in the relevant field of knowledge, and secondly, its statistical consistency should be demonstrated in order to reliably measure the degree of skill acquisition. Teachers should review the relevance of the items, which should be measured on a Likert scale, while an indicator such as Cronbach's Alpha can provide statistical certainty of consistency to the rubric that is used.

This article discusses how the evaluation system can be improved through statistical analysis of correlation between the criteria and parameters that make up the structured evaluation guide or rubric. The analysis was applied to a specific case within the Industrial Engineering career, for a government institution, in the State of Mexico, with the aim of strengthening sustainability in the training of engineers. The improvement process involved increasing consistency by strengthening the criteria or subtests that had a low correlation and were susceptible to restructuring. This approach can lead to a more effective evaluation system, which in turn can contribute to better training outcomes for students in the field of engineering.

As a result of the improvement process discussed in the previous section, the Cronbach's Alpha index of the rubric, which originally did not show an appropriate level of consistency, was increased to a normally accepted level, providing a greater degree of statistical reliability. With this increase in consistency, the rubric can now be used as an effective evaluation instrument, presenting the necessary statistical consistency and reliability for use in the competency evaluation process. This example can be used as a model to improve other rubrics, leading to more reliable instruments that can improve the evaluation of skills in any teaching-learning discipline.

In order to fully realize the benefits of rubric-based evaluation systems, it is crucial that the teachers who will be using the rubric have adequate training on how to interpret the work that will be evaluated. This will help ensure that the evaluation system does not generate grades that have significant dispersion in the competency evaluation of engineering programs. By providing training on how to use the rubric, teachers can make more informed and accurate evaluations, leading to a more effective and reliable evaluation process. This training can also help to promote a more consistent understanding of the evaluation criteria among instructors, which can further contribute to the overall effectiveness of the evaluation system.

Acknowledgment

The authors express their gratitude for the support provided by the National Technology Institute of Mexico, TES Cuautitlán Izcalli, which made it possible to complete this research work and share the knowledge at this event.

References

[1] ISO Quality management systems — Requirements. 2015. Switzerland: OBP. https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:es
[2] IATF 16949 - Customer Specific Requirements. 2016. https://www.iatfglobaloversight.org/wp/wp-content/uploads/2019/05/IATF-16949-GM-CSR-May-2019.pdf
[3] B. Jackel, J. Pearce, J., A. Radloff, and D. Edwards. "Assessment and Feedback in Higher Education: A Review of Literature for the Higher Education Academy". *Higher Education Academy*. 2017.
[4] M. Bearman and R. Ajjawi. "From "Seeing Through" to "Seeing With": Assessment Criteria and the Myths of Transparency". Frontiers, 3, 2018.
[5] Nadolski, R., Hummel, H., Rusman, E., et al. "Rubric formats for the formative assessment of oral presentation skills acquisition in secondary education*". Education Tech Research Dev* 69, 2663–2682. 2021.
[6] J. Tejada and C. Ruiz (2016). "Evaluation of professional competencies in higher education: challenges and implications". *Education XX1*, 19, (1), 17 – 37, 2016.
[7] Brookhart, S. "How to Create and Use Rubrics for Formative Assessment and Grading". Alexandria, VA: ASCD, 2013.
[8] Andrade, H. "Using rubrics to promote thinking and learning". *Educational Leadership* 57, 13–18, 2000.
[9] J. Arter, and J. Chappuis. "Creating and Recognizing Quality Rubrics". Boston: Pearson, 2007.
[10] S. Brookhart and F. Chen. "The quality and effectiveness of descriptive rubrics". *Educational Review*, 343 - 348, 2014.
[11] E. Rusman and K. Dirkx. "Developing Rubrics to Assess Complex (Generic) Skills in the Classroom: How to Distinguish Skills' Mastery Levels?". *Practical Assessment, Research & Evaluation*, Vol 22 (12), 2017.
[12] R. Szafran. "The Miscalculation of Interrater Reliability: A Case Study Involving the AAC&U VALUE Rubrics". *Practical Assessment, Research & Evaluation*, 22(11), 2017.
[13] S. Brookhart. "Appropriate Criteria: Key to Effective Rubrics". *Frontiers*. Educational Review, Sec. Assessment, Testing and Applied Measurement, Volume 3, 2018.
[14] S. Brookhart, and F. Chen. "The quality and effectiveness of descriptive rubrics". *Educational Review*, 343 - 348, 2014.
[15] J. Bernal, M. Bernal, and M. Monroy, M. "Measurement System Analysis applied to a competence assessment methodology in engineering training programs in Mexico". *LACCEI Proceedings*, Buenos Aires, Argentina, 2023.
[16] Minitab. "Analysis of elements. What is internal consistency?" Technical support Minitab 20, 2023.
[17] R. Hernandez and C. Mendoza. Investigation methodology. Mexico: McGraw Hill, 2023.
[18] J. Norcini. "Standards and Reliability in Evaluation: When Rules of Thumb Don't Apply." *Academic Medicine* 74 (10): 1088–1090, 1999.
[19] D. Iacobucci and A. Duhachek. "Advancing Alpha: Measuring Reliability with Confidence." *Journal of Consumer Psychology* 13 (4): 478–487, 2003.
[20] L. Cronbach. Coefficient alpha and the internal structure of tests. Psychometrika, 16 (3), 297 – 334, 1951.
[21] J. Del Pozo. Professional skills. Evaluation tools: the portfolio, the rubric and situational tests. Spain: Narcea, 2013.
[22] S. Morales, R. Hershberger, and E. Acosta. "Evaluation by competencies: how is it done?" *Medical School Magazine*. Vol. 63 (3), 2020.
[23] M. Young, *The Technical Writer's Handbook*, Mill Valley, CA: University Science, 1989.