COMPETENCY-BASED ASSESSMENT AND THE USE OF PERFORMANCE-BASED EVALUATION RUBRICS IN HIGHER EDUCATION: CHALLENGES TOWARDS THE NEXT DECADE

Katherina Gallardo
Monterrey Institute of Technology and Higher Education, Mexico
E-mail: katherina.gallardo@tec.mx

Abstract

Rubrics are assessment guides for grading and giving feedback to students while demonstrating acquired knowledge and skills. In the last decades, rubrics have been all part of the most used learning evaluation tools in higher education. Its use is further well-related to competency-based assessment purposes. Nevertheless, criticism around design and application has been also implemented in certain reports. In order to understand rubrics' evolution, practice, benefits, and trends, a systematic literature review on rubrics' design and use has been conducted. Two databases were selected: Scopus and ProQuest Education. Two phases were determined: The first allowed to identify articles related to rubrics' design and application for almost three decades. 584 articles were found. From these, most cited articles served to give a scope of rubric evolution and trends. The second phase permitted to identify design and use of performance-based evaluation rubrics from 2009 and 2019. Five terms and Boolean combinations were used. A total of 259 articles was found. After analyzing abstracts and content, 11 open access articles on performance-oriented rubric design and application were chosen. As a result, some facts and reflections on rubric design complexity are discussed towards responding to educational challenges related to competency-based assessment contemporary demands: integration of human learning domains going beyond cognition, authenticity, and interdisciplinarity as features that characterize learning design situations, follow students' progression, educator's preparation on assessment literacy for responding to CBA demands, and the involvement of experts of the work field in determining essential evaluation indicators as the main topics looking forward to the next decade.
Keywords: *competency-based assessment, competency-based education, higher education, learning taxonomy, rubric design.*

Introduction

Competency-based assessment (CBA) has been a relevant educational topic for almost 25 years. It is common knowledge that its complex condition comes from its roots in competency-based education (CBE) as an integrated paradigm derived from several pedagogical approaches. Therefore, it is not possible to refer to a single entity, but rather, in practice, there are multiple ways to define and operationalize it (Farris, Demb, Janke, Kelley, & Scott, 2009; Gervais, 2016; Meyer-Adams, Potts, Koob, Dorsey, & Rosales, 2011; Rodriguez & Gallardo, 2017).

The International Bureau of Education of UNESCO emphasizes the complex outcomes of a learning process that implies working within CBE, as knowledge, skills, and attitudes are all part of the learning process, rather than focusing on traditionally defined subject content. Thus, learners are expected to acquire and apply knowledge, skills, and attitudes to perform in

62

situations they encounter in everyday life (Escudero Escorza, 2016; Kuh, Jankowski, Ikenberry, & Kinzie, 2014). This concern has been analyzed several times by educational authorities, in order to fulfill the needs of governments, educational institutions, and teachers for planning, implementing, and assessing performance. In fact, they are obliged to render counts to society about how close students are in reaching the standards of stated competencies. Then, CBA becomes fundamental for ensuring educational quality and accountability (Duran, Estay-Niculcar, Alvarez, & Randolph, 2015; Koh, Tan, & Ng, 2012).

CBA is based mainly on performance and authentic assessment. It has been defined as involving observation and assessment of students' behavior while making things (Gao & Grisham-Brown, 2011; Hancock, 2007). Students are asked to demonstrate that they are capable of transferring knowledge, executing some skills in a certain complex situation in order to solve problems, and taking decisions, or creating new alternatives within a given context. This performance is generally observed in a semi-structured context or in real-life situations where collaborative work is required. Some advantages of applying performance and authentic assessment are: (a) giving more meaning to complex learning targets; (b) assessing students' ability to make decisions and solve problems; (c) integrating knowledge, skills, and attitudes into assessment and feedback; (d) giving the learning processes an important role rather than only focusing on solutions and products; and (e) being consistent with philosophical and pedagogical CBE basics.

Most of the time, discussions about implementing CBA focus on how to guarantee that performance assessment would be done with feasibility and validity, without depending just on standardized tests (Barber & Phillips, 2000; Duckworth & Yeager, 2016). Instead, alternative and authentic assessment practices are required (Caliskan & Kasikci, 2010). Thus, some alternative assessment practices emerge and are used by teachers: project-based learning, challenge-based learning, portfolio, etc. Nevertheless, poor/fast teacher training and several misunderstandings while designing and interpreting data through these assessment mechanisms could be considered weak areas of CBA implementation.

Competency-based Assessment: Challenges in Times of Educational Change

Almost all societies are living in a period of rapid and disruptive educational development. Having access to a high-quality education that considers human beings' main concerns and needs of development besides simple disciplinary knowledge, becomes part of worldwide priorities to be solved in the future. European, Asian, Australian and Latin-American countries are modifying their educational scope structurally in order to introduce CBE at different levels: primary, secondary, vocational and tertiary education (Boukhentache, 2016; Voronov, Chasovskykh, Popov, Anyanova, & Krainova, 2015).

Even if there has been a high interest in implementing CBE over these years as one possible way to give answers to the Economy and Society Knowledge demands (Carrillo, 2015; Slaus, 2013), there are aspects that are not totally solved around how to assess students' performance with accuracy and feasibility, in a world that has explained learning results based simply on grades and statistics for years. Indeed, identifying and recognizing CBA advantages does not automatically change traditional learning assessment practices. However, we cannot deny that the scope is changing progressively.

For instance, the Australian Learning and Teaching Council has made some relevant assessment decisions (Boud, 2009) related to strengthen CBA toward the year 2020: (a) assessment is a central feature of teaching and the curriculum; (b) it is an accurate way to understand how students learn and what students achieve; and (c) it is one of the most significant influences on students' experience, so it is mandatory to improve assessment practices due to its huge impact on learning quality.

Another example of the CBA evolution is taking place in Bangladesh (Roy, 2016). The Ministry of Education of this country is looking forward and moving faster to implement new ways to improve the quality of basic education by evaluating learning achievement in terms of terminal competencies, considering different school systems based on a nationally representative sample, and examining the relationship between learning achievement and selected students' characteristics, such as socioeconomic status, school-related variables, and receipt of extra educational inputs. Thus, the government is applying different assessment strategies rather than just examining and scoring content retention through standardized tests.

Moreover, it is understood that CBA not only has challenges to face and solve in the implementation phase but also a number of issues to address for gaining credibility (Barber & Phillips, 2000; Ellis et al., 2015). Among these main challenges, the most relevant are:

- Revising competencies and academic programs to make sure CBA is conducted with coherence
- Deciding processes, terms, and parameters related to competence accomplishment along with each program
- Counting on internal and external experts' support for establishing valuable tasks and projects
- Establishing collegiate agreements for the performance-based evaluation rubric
- Selecting meaningful standards for assessing and giving feedback.

Thus, it can be affirmed that in the eye of this hurricane of educational changes toward adopting CBA, big efforts must be made to deliver a high-quality assessment process.

Rubric: More than 20 Years of Research

The operationalization of CBA into applied strategies and the use of different tools for observing, collecting, and giving feedback about students' performance have been studied and enriched over the past few years. Undoubtedly, one of the most-used performance-assessment tools is rubric. Its main intention is to describe what students are expected to demonstrate through their performance. This is possible as long as a rubric contains a clear description of intended and connected actions corresponding to the statement of criteria, and the standards and levels of performance expected from the related competencies (Jonsson, 2014).

Rubrics for learning assessment purposes have been studied for several years. From the first discussion about its use within holistic evaluation processes (Grubb, 1981), almost six hundred articles have been published to discuss new knowledge on the design or use of this tool in different disciplines (Elsevier, 2019). Thus, rubric has passed through a rapid evolution process over the past few years (Bharuthram & Patel, 2017).

Rubric design has challenged educators and educational researchers to study, understand and create different ways to answer one of the most difficult aspects of the learning process: giving judgments about performance. These judgments can come from different actors involved in the educational process, even from experts in the field of study who are asked to give their opinion about students' performance. Then, rubric is expected to be a simple, clear, accurate, consistent and flexible tool for collecting diverse points of view related to certain performance criteria, in a holistic or an analytic way (Mertler, 2001; Panadero & Jonsson, 2013).

Many efforts have been made lately to establish norms to improve rubrics' design and use from the technical point of view. Curiously, this norming trend comes from engineering education initiatives. For instance, Goldberg (2014) discussed some common problems while designing a rubric: lack of consistency and parallelism, detection of "orphan" and "widow" words and phrases, redundancy in descriptors, inconsistency in focus of qualifiers, limited routes to partial credits, unevenness in incremental levels of performance and inconsistencies across the suite of rubrics. Additionally, Schoepp, Researcher, and Kranov (2018) established

64

a norm for designing accurate and valid rubrics focusing on a higher education context, which consisted of document preparation; rubric review; initial reading and scoring of one learning outcome; initial sharing/recording of results; initial consensus development and adjusting of results; initial reading and scoring of remaining learning outcomes; the reading and scoring of remaining transcripts; sharing/recording results; and the development of consensus and adjustment of results.

Nevertheless, in the literature, profound discussions about the need for alignment among educational model features are rarely found, nor are performance-assessment principles and rubric design parameters, especially in terms of CBA.

Research Problem

This research intention was to have a wide perspective on one of the most-used performance-assessment tools: the rubric. It has become a powerful way to assess and give feedback to students in the past few years (Jonsson, 2014; Lasater, 2006; Reddy & Andrade, 2010). Nevertheless, several positive as well as negative ideas have been discussed about its design and use in different disciplines. It could be relevant to understand practices, weak areas as well as challenges for the next years in terms of its use. The two researched questions that guided this research were: What can it be said and understood about rubric evolution in the last decades? and What are the main challenges about performance-based evaluation rubric design and application?

Research Methodology

General Background

The aim was to understand rubrics' design and use in tertiary education in general and for performance assessment purposes in particular while CBE model is getting used and expanded worldwide. This aim involved a Systematic Literature Review (SLR). This methodology aims to identify, evaluate and summarize the findings of all relevant individual studies, thereby making available evidence more accessible. The process includes elements such as (1) start by a research question; (2) specify the search strategy and criteria for inclusion and exclusion of studies; (3) identify, describe and synthesize relevant research. The methodology enables the incorporation of quantitative, qualitative, and mixed methods studies as long as the importance of both the intellectual quality of the original studies and the context, culture, and values relevant in evaluation studies are included (Caracelli & Cooksy, 2013).

Two abstract and citation database of peer-reviewed literature databases were selected: Scopus and ProQuest Education. For the searching process five keywords were chosen according to the research interest: Rubrics, Competency-based education, Competency-based assessment, Performance Assessment, and Competence. It was determined that "and" was the right Boolean command for this search purpose. The sample then was determined by the search process run on these databases combining main and subsequence keywords.

Instrument and Procedures

The researcher counted on access to make searches in two databases through the university digital library system (https://biblioteca.tec.mx/basesdedatos). Hereby, the phases and steps followed for the analysis.

Phase I: (1) Enter to the databases; (2) use of the keyword "Rubric" in the title of the

Katherina GALLARDO. Competency-based assessment and the use of performance-based evaluation rubrics in higher education: Challenges towards the next decade

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Vol. 78, No. 1, 2020

65

article; (3) select a period of 23 years (1996-2019) for search purposes, as 1996 was identified as the year where the level of scientific production in this topic started to increase.

Phase II: (1) Enter to the databases; (2) use of five selected words for the search purpose: Rubric as the main word in the title of the article and Competence, Competency-based education, Competency- based assessment and Performance assessment as complimentary words in the title, keywords and abstract. The Boolean combination "and" was used for filter of search: Rubric and Competency-based education in the title, abstract and keywords, Rubric and Competence in the title, Rubric and Competence-based assessment in the title, abstract and keywords, Rubric and Competence-based assessment in the title, Rubric and Performance-assessment in the title, abstract and keywords and Rubric and Performance-assessment in the title identify; (3) gather systematically all the findings in independent datasheets for making searches and select specifically the articles corresponding to research purpose.

Data Analysis

The database was organized in an Excel file with a total of 11 independent sheets. In each sheet, the list of collected data was registered. The headings for storing systematically each article data were: Authors, Author(s) ID, title, year, source title, volume issue, article number, page start, page end, page count, cited by, DOI, link, affiliations, abstract, author keywords, document type, publication stage, access type. These fields were analyzed for ensuring that any article was duplicated. A counting table was designed in a complimentary independent sheet for knowing exactly how many articles were found in this first moment (Table 1).

Table 1
SLR strategy and criteria for inclusion and exclusion of studies

Phases of research	Database	Main word	Complimentary words	Amount
Phase I	Scopus & ProQuest	Rubric (TIT)	D/A	584
Phase II	Scopus	Rubric (TIT)	Other four keywords	153
	ProQuest	Rubric (TIT)	Other four keywords	112
	Scopus & ProQuest	Rubric (TIT)	Other four keywords	*259
	Scopus	Rubric (TIT)	Competency-based education (TKA)	9
	ProQuest	Rubric (TIT)	Competency-based education (TKA)	7
	Scopus	Rubric (TIT)	Competence (TIT)	10
	ProQuest	Rubric (TIT)	Competence (TIT)	2
	Scopus	Rubric (TIT)	Competency-based assessment (TKA)	6
	ProQuest	Rubric (TIT)	Competency-based assessment (TKA)	7
	Scopus	Rubric (TIT)	Performance assessment (TKA)	141
	ProQuest	Rubric (TIT)	Performance assessment (TKA)	103
	Scopus	Rubric (TIT)	Performance assessment (TIT)	20
	ProQuest	Rubric (TIT)	Performance assessment (TIT)	3

Note: TIT means title, TKA means title, keywords and abstract, D/A means does not apply
* 6 articles were deleted: 3 duplicates, and 3 products not recognized as research articles

Research Results

Phase I Results

The first analysis of the 584 collected articles related to rubric design and application from the title of these works, permitted a first glance of their origins: these articles have been published mainly in these countries: United States (31%), Spain (12%), Australia (10%), Canada (8%) and Turkey (7%), United Kingdom (6%), Netherlands (5%) Japan (5%). Other countries have 2% or less. Besides, faculty highly involved in the research of this topic are affiliated to Universidad Autónoma de Madrid, Purdue University, University Albany State of New York, Universidad Jaume and Texas A&M University.

From the 584 articles, 29 were chosen by their high number of citations as a criterion that might take the researcher to read the most popular publications in this subject area along last years. This permitted to revise carefully the content, orientation, results, and discussions around rubric application.

In this phase of the research, it was also possible to map the different disciplines involved in the study of rubric design and application besides social science and education. It was interesting to find that many articles were published by researchers of other disciplines different from social studies, such as engineering, health science, business, agriculture, etc.

Besides, it was detected first that most of the articles emphasize the advantages that rubrics bring to diagnostic, formative, and summative assessment, specifically on certain features such as utility and transparency while rubric is used for giving feedback. Besides, its power of making students reflect around self-assessment and self-efficacy practices is being discussed (O'Brien, Franks, & Stowe, 2008; Smit, Bachmann, Blum, Birri, & Hess, 2017; Yune, Lee, Im, Kam, & Baek, 2018).

Nevertheless, some misunderstandings and criticism around rubric design process are also debated. For instance, Tigelaar and Janssen (2012) reported some dilemmas that arose while designing rubrics for assessing science student teachers' performance: fully detailed criteria against open-ended description; the importance given to the theoretical domain against practical knowledge application; the establishment of novice against expert performance level as point of departure for performance explanation; the involvement for validation purposes of only teachers against including some stakeholders, among others. Moreover, Eshun and Osei-Poku (2013) and Jonsson and Svingby (2007) highlighted some disadvantages and criticisms of designing and using rubrics: the exclusive and excessive use of rubrics for assessment purposes, a lack of guidance for making judgments through rubric levels, the absence of a validity and reliability process, were among the main problems identified. Table 2 contains an overview and summary of Phase I synthesis of rubric research and evolution in the last decades.

Table 2

Rubric evolution based on a Scopus and ProQuest analysis of articles about rubric design and application over the past 23 years

	1996-2000	2001-2005	2006-2010	2011-2015	2016-2019
Number of scientific published articles	23	32	79	188	262
Disciplines besides Social Science and Education	Pharmacology Physics Veterinary	Mathematics Medicine Nursery Neuroscience	Earth Science Economics Energy Engineering	Chemistry Computer science Dentistry	Agriculture Biochemistry Business Creative thinking
Parallel emergent topics related to rubric design and application	Formative assessment Feedback	Validity Reliability Quantification analysis Scoring Design recommendations	Self-assessment Self-efficacy Teaching Inter-rater and intra-rater reliability	Higher education Peer assessment Reasoning skills Writing skills Factor analysis	Metacognition
Criticism	Risk of exclusive use	Lack of validity and reliability Right number of criteria	Time consuming design and use Intolerance to change Converting scales to scores	Redundancy Inconsistency in focus qualifiers	Students lack of participation in design

Phase II Results

From the 259 articles, 11 open access research articles were chosen for a deeper analysis according to SLR methodology. Their contents have been considered of high involvement with performance-based evaluation design and application rubric topic. First, a summary of main aspects of each article was organized in a table. Then, three questions were asked to extract relevant information about performance-oriented rubric research: (1) what are the main interests while studying performance-oriented rubric design and use? (2) what kind of data are these studies collecting for analysis; (3) what is the proposal for future research? Finally, some facts, as well as some missing aspects according to competency-based assessment and performance-assessment demands, are discussed.

The main aspects related to the 11 chosen articles were collected. It was taking into consideration: Author(s) and year of publication, the academic level where the research was conducted, the subject area or discipline, focus of the research problem, methodology, main results and suggested future research. The information is organized in table 3 following a chronological order.

68

Table 3
Summary of main facts and findings from 11 open access articles chosen for SLR

Author(s) and year	Academic level	Subject area	Focus of the research problem	Methodology	Results	Future research
Latimer, Bergee, and Cohen (2010)	High school – large groups festival	Music performance assessment	Analyze reliability of rubrics as well as the level of acceptability of use.	Quantitative and qualitative: (1) an ex-post-facto study by collecting 2016 results from rubric application for estimating music students' performance in music festivals along 3 consecutive years. (2) categorizing comments collected from 515 open-ended questionnaires to estimate utility and improvement suggestions.	Positive: moderate internal consistency or the rubric (0.88) and moderate reliability at all levels (by dimension and total). These results were similar to what other studies in this field have published. Utility level was also positive according to rubric adjudicators.	Reinforce the practice of using rubrics for music performance assessment.
Kan and Bulut (2014)	8th grade	Teachers' use of performance-oriented rubric	How does the rubric use influence the scoring of performance assessment? Is there any interaction between teachers' work experience and their grading behaviors?	Quantitative: Crossed random-effects modeling to estimate rater effects among teachers and effect of teaching experience. Three facets were established: persons, tasks, and raters. 50 students 17 teachers.	Positive: The necessity of using scoring a rubric for grading performance assessment is proved.	Study other types of rubrics for performance assessment purposes
Verano-Tacoronte, González-Betancor, Bolívar-Cruz, Fernández-Monroy, and Galván-Sánchez (2016)	Higher education-Undergraduate	Students' oral skills	Analyze the reliability and validity of a rubric designed to guide and evaluate oral presentation skills, with support in visual media and with a time limitation, in a work context in team.	Quantitative: Pearson correlation for estimating global scores giving by two professors. Then, Cohen Kappa score for estimating the level of agreement.	Positive: The results revealed that there was a relatively high positive correlation among the score of each evaluator as well as an acceptable level of consensus One of the factors that contributed to the acceptance of the rubric was the previous training of the appraisers through of different actions.	Make more research including the evaluation of the standard of utility

Katherina GALLARDO. Competency-based assessment and the use of performance-based evaluation rubrics in higher education:
Challenges towards the next decade

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Vol. 78, No. 1, 2020

69

Reference	Level	Field	Objective	Methodology	Results	Recommendations
Jin, Song, Shin, and Shin (2015)	Higher education Undergraduate	Engineering design	Develop a performance-based evaluation rubric that can assess and enhance students' engineering design skills in introductory engineering design courses	Quantitative mainly. The study was divided into five moments: (1) considerations about engineering design phases: problem, solution and implementation; (2) establishment of performance criteria and tasks; (3) and decisions about scoring scales: 1: poor; 2: marginal; 3: satisfactory; 4: excellent; and (4) validation review by experts using three methods: expert review, a usability test, and a field evaluation. (5) Students 'perception after its use.	Positive: According to the expert review, the average score was 3.65 (out of 4.0), and all content validity index and inter-relater agreement values for each item exceeded 0.8. These results indicate all elements to be highly valid and reliable. According to the students' perception of the learning the rubric was helpful for student activities and could be used as a useful guide for successful engineering design activities.	Not included
Crespo-Martín, Figueras-Ferrer, Mascarella-Vilageliu, Nualart-Torroja, and Valle-Martí (2016)	Higher education-undergraduate	Soft skills (teamwork)	Design a model of evaluation headings *ad hoc*, and to carry out a pilot test to verify its effectiveness, evaluate the results obtained and evaluate the results.	Qualitative. The study was divided into two phases: (1) orientation and support sessions for professors to get to a consensus about expected achievement (2) the definition of the competencies, table of headings, implementation of the rubric and analysis of the results.	Positive: (1) training work highlighted lack of consensus. Then putting definitions in common served to get to a rubric with a better structure. (2) getting to an agreement on the definition of teamwork competence. (3) the pilot experiences contribute to make deeper reflection on the competence and way to develop it.	Establish a department or center to consolidate all performance assessment procedures

https://doi.org/10.33225/pec/20.78.61

70

| Brown (2017) | Higher education - Undergraduate | Teachers' assessment competencies | Make an analysis of the components of a student teaching practicum rubric used in teacher preparation program to determine if the performance-based assessment supported the expectations and requirements of a teacher performance assessment. | Qualitative: several information sources like planning, observation criteria, instruction, standards and assessment practices were all analyzed. | Positive: Although rubrics required an initial investment of time, aligning observation rubrics with the requirements of teacher performance assessments provided teacher candidates a valuable information source about expectations of the examinations and assessments required for initial teacher certification. This information provided them the opportunity to analyze their own work and progress towards to fulfill the expectations of the teacher performance assessment. | Continuously examine how faculty introduce professional teaching standards, performance-based assessment and evaluation instruments in the teacher preparation program. |

Lee and Fields (2017)	Higher education - Undergraduate	Makers movement and electronic textiles	Characterize competence change in undergraduate students in a semester using e-textiles. The focus was the rubric and its illustration in each dimension rather than describing all three areas of learning simultaneously. These three dimensions were: Circuitry, crafting and computation.	Qualitative: pre and post interviews with 7 students. A coding scheme was developed and organized into an analytical rubric to understand performance in the three dimensions	Positive: Tracking students' way of learning through the interviews permitted to design a a detailed way to map learning in different areas of e-textile design than those examined before. Besides, as the rubric was developed from student interviews within the situated context of a specific e-textiles course. It can provide a starting point for investigations of other e-textiles learning environments, such as those beginning to reach into K12 classrooms where assessments are being developed.	Continue making research about learning with other e-textile student groups
Carson and Kavish (2018)	Higher education - Undergraduate	Writing skills Sociology studies	Explore if the use of feedback comments, a graded rubric and scaffolding within a discipline will increase writing skills	Quantitative: eight writing scaffolding groups were divided into four, from less complex to more complex writing skills goals to achieve. Bloom's taxonomy was used to operationalize progression along writing skills performance improvement.	Positive: The use of rubric based on Bloom's taxonomy and feedback were useful to follow the performance track along the four writing courses. The initial writing skills could vary based on the quality of writing instruction students receive while in high school. However, it is reasonable to assume that the scaffolding of rubrics would be beneficial to all students, regardless of the initial writing skill of students when they begin their higher education.	Not included

72

Hindriana and Setiawati (2018)	Higher education – Undergraduate	Biological education	This research had two focus: the first one to understand the pattern of development of authentic assessment rubric, gather accurate information related to students' performance and learning achievement. The second one to designing authentic assessment rubric. The research started it by deciding the aims, and components needed to be in the rubric.	Quantitative: 97 students participated in the study. It was divided into two phases: (1) Decide the components in authentic assessment design by: a) making requirements; b) deciding on authentic assessment that can develop students' potential in hard skill and soft skill; c) making indicator in cognitive, psychomotor, and affective domains; d) deciding strategy of development and application in the class. (2) Implementation on the ongoing assessment and doing research in time series to observe the development of students' potential along with the use of authentic assessment rubric.	Positive: Validity and reliability of problem solving and journal analysis skill rubric Had a big loading factor: above 0.50. Thus, most of the items were valid. Reliability stated as adequate: Cronbach greater than or equal to 0.70	Not included
Kopp (2018)	Higher education – Undergraduate	Health and clinical instructors	Test the hypothesis: a reliable assessment method will detect increased scores from midterm to final evaluations. Then, reliability of an undergraduate clinical performance grading rubric was the focus of this research study.	Quantitative: Rubrics' reliability and consistency were measured through written assignments and clinic performance observations. 58 first semester students Independent sample T test and Pearson correlation were used. Cronbach's alpha was used to measure rubric consistency outcomes.	Positive: Significant difference was found between midterm and final performance evaluations. No correlation was found between written assessment and final performance evaluation. Rubric consistency was high (0.91)	Replicate this study in multiple health schools including interrater assessment.

| (Velasco-Martínez & Hurtado, 2018) | Higher education-Undergraduate | Rubric design analysis | Make an analysis of 59 rubrics designed by different professors in three different universities in Spain. | Both, quantitative and qualitative: Exploratory study. The analysis included the estimated frequencies of: (1) Type of tasks evaluated: written products, oral presentations, use of software, problem solution, teamwork, simulations, audiovisual design, and others specific from certain disciplines. (2) type of competences: instrumental, interpersonal, systemic, combination of skills (3) different disciplines: social science, engineering, health, science, arts and humanities. | Both, positive and negative. Positive: Criteria homogeneity, appropriate progression along levels, language use parallelism. Negative: Lack of explanation and poor level of specifications about criteria and purposes, combination of scales types. | Continue exploring ways as well as training professors to evaluate students' performance |

What are the main interests while studying performance-based evaluation rubric design and use? It could be affirmed that there are two main research interests: In the first place, reliability and validity as a way to ensure rubric quality. The goal refers, on the one hand to consistency and stability of the instrument. On the other hand, to the consensus from experts to select criteria, describe expected performance levels and agree upon meaning. Indeed, this interest has been considered relevant since 2001 (Table 1). It is possible to infer from Latimer, Bergee and Cohen (2010), Jin, Song, Shin and Shin (2015) and Kopp (2018) that performance-based evaluation rubrics are instruments with high possibilities on supporting reliable and valid assessment process for observing and judging students' performance advances.

In the second place, the capabilities and preparation of educators for designing rubrics. Undoubtedly, rubric design is a complex task that involves decision making starting by the understanding on student's profile, ways to promote progression along the educational program, the context and conditions of the educational scenario, the task purpose, previously defined competencies and standards, authenticity of tasks, among other variables. Velasco-Martínez & Tójar Hurtado (2018) concluded after a wide analysis of diverse rubrics for different disciplines assessing needs that there is a latent risk educator that could have designed rubrics without any previous training. Thus, it is recommended to establish explicit and clear criteria to ensure rubrics design quality. Therefore, it seems advisable from the analysis that participation in workshops of design and use of rubrics in order to train teachers for an adequate evaluation with rubrics is highly recommended.

What kind of data are these studies collecting for analysis? From the 11 articles analyzed, almost all of them presented quantitative data analysis. This information was used for demonstrating through conventional statistics the level of reliability and correlation of rubrics results in contrast to grading. Cronbach's alpha and Pearson correlation coefficient were the most used measures to make judgments about rubrics quality.

Nevertheless, results from studies based on mixed methods or qualitative methodology gave information about the design process and perceptions of users rather than results of its application. These studies focused on decision making, consensus, and proposed different ways of making rubrics useful. For instance, in the study of Brown (2017) as a proposal for improving performance-based evaluation rubric design started by the examination of certain documents and procedures that guide educators' performance. In the case of Lee and Fields' study (2017) the main information source for rubric design was students' perception. Neither theoretical contents, nor teachers' expectations but the way students perceived their performance advances in three different tasks along a semester were taken into consideration.

What is the proposal for future research? Most researchers' point of view agreed that the application rubrics in other scenarios as well as the exploration of new proposals for rubric design are the most common recommendations. Besides, educators training and practice for designing and improving rubrics should be also considered for future research.

In the opinion of the researcher, it was unexpected to find just a very few ideas for future research while performance-based evaluation rubrics is a complex vast topic with a wide spectrum of educational research possibilities. This situation derived to certain aspects of discussion.

Discussion

Results from this research permitted to confirm that rubrics design is a topic of upward interest in the educational process around the world. The rise of interest from decade to decade doubled and tripled the number of studies published in recognized indexed journals. Thus, SLR permitted to have a holistic view of rubric evolution for almost three decades as well as a current view around relevant issues in the last decade about its design, application and quality insurance for performance-assessment purposes.

Katherina GALLARDO. Competency-based assessment and the use of performance-based evaluation rubrics in higher education: Challenges towards the next decade

PROBLEMS
OF EDUCATION
IN THE 21ˢᵗ CENTURY
Vol. 78, No. 1, 2020

75

One of the main results from the SLR is the confirmation of the importance of the design and use of rubrics as a matter of studies in different disciplines' educational processes as well as soft skills assessment interests. Undoubtedly, the popularization of rubrics as a tool for guiding, observing and measure students' progress in different knowledge areas and task is increasing.

Besides, it was confirmed through the specific analysis in Phase 2 that higher education is the education level where most of these studies are taking place. Nevertheless, the advances in rubric studies related to complementary topics of CBA and performance assessment are quite weak at this time. Indeed, just a few isolated efforts could be found through SLR that give a deeper understanding on how to integrate in rubrics design elements of authenticity and complexity of learning scenarios (Hindriana & Setiawati, 2018), full integration of other domains going further cognition (Lee & Fields, 2017) or assessment literacy focused on performance-based evaluation rubrics design (Velasco-Martínez & Hurtado, 2018).

As a matter of fact, there are some educational challenges around CBA and performance assessment that educators, as well as educational researchers, would need to identify for the next decade related to the study of CBA and performance-based evaluation rubrics:

(1) Integration of cognitive, psychomotor, social and affective domains. Even if almost all studies refer to performance-based evaluation rubrics as powerful tools to observer capabilities and executing some skills, very few of them go beyond integrating elements distinct from the cognitive domain. This fact could be considered contradictory as psychomotor, social and affective domains need to be integrated into students' performance in complex situations based on competencies development (Escudero Escorza, 2016; Kuh et al., 2014).

(2) Comprehension of the authentic complex situation where performance assessment takes place. Authentic complex situations (Gao & Grisham-Brown, 2011; Hancock, 2007) need to be comprehended to clarify the indicators that integrate a performance-based evaluation rubric. This is even more important when problems or situations to be solved integrate interdisciplinary activities. As one of the most important performance assessment goals is giving information on the process of gaining expertise, then the identification and operationalization of the main variables and circumstances that configure a complex situation to be solved are critical pieces of information for learners and educators.

(3) Definition of expected performance reached level according to students' progress. From the 11 studies, just Brown (2017) and Lee & Fields (2017) took explicit consideration of this variable related to one of CBA main purpose. Indeed, performance-based evaluation rubrics should permit to understand the way students progress from one time to another. This is possible when some consensus about the expectations on students' performance level is previously established.

(4) Assessment literacy in educator's preparation for CBA demands. From the 11 studies, just one of them (Velasco-Martínez & Hurtado, 2018) focused on the understanding of how educators, specifically professors, design rubrics. The report revealed among the most relevant weakness aspects: the lack of explanation, poor level of specifications about criteria and purposes, and possible misunderstandings in the selection of scales. As educators are the main responsible for instrumental assessment practices in the classroom, it would be necessary to identify good practices while taking decisions for rubric design purpose.

(5) Participation of external experts in rubric design. CBA features of authenticity involve challenging situations from the work field (Gao & Grisham-Brown, 2011; Hancock, 2007). Students are exposed intentionally to these situations to gain expertise and improve their level of performance in certain competencies. Then, it would be appropriate to take opinions from experts of diverse disciplines as

76

employers and seniors while designing a performance-based evaluation rubric as well as when applying and giving feedback to students.

Conclusions

Rubrics have evolved in the last years as a powerful support tool to make judgments about students' learning in several disciplines. The amount of studies related to rubric design, application and reliability for ensuring high level of quality and utility has increased globally. The frequency of the studies related to this evaluation tool is concentrated by now on higher education mainly.

Rubric design features have been changing according to the educational model and assessment needs. In the case of this study, performance-based evaluation rubrics have been considered a relevant topic as CBE and CBA have transformed higher formative education scope in many countries around the world.

In the framework of CBA and performance assessment settings, a SLR took the researcher to focus on 11 articles related to the design and use of performance-based evaluation rubric. The analysis reflected that almost the same research topics are also studied, even if CBE scope would have conducted researchers to incorporate other subjects such as integration of human learning domains going beyond cognition, authenticity and interdisciplinarity as features that characterize learning design situations, follow students' progression along their higher education preparation, the involvement of experts of the work field in determining essential evaluation indicators or assessment literacy in educator's preparation for CBA demands. These aspects need to be considered in the next decade of performance-based evaluation rubrics studies.

Acknowledgements

References

Barber, M., & Phillips, V. (2000). Big change questions: Should large-scale assessment be used for accountability? *Journal of Educational Change, 1*(3), 277–281. http://dx.doi.org/10.1023/A:1010064308012

Bharuthram, S., & Patel, M. (2017). Co-constructing a rubric checklist with first year university students: A self-assessment tool. *Journal of Applied Language Studies*, 11(4), 35–55. http://dx.doi.org/10.17011/apples/urn.201708073430

Boud, D. (2009). Assessment 2020: Seven propositions for assessment reform in higher education. *Education and Training*, 330–333. https://doi.org/10.1200/JOP.2011.000236

Boukhentache, S. (2016). Operationalization of competency-based approach: From competency-based education to integration pedagogy. *Arab World English Journal, 7*(4), 439–452. https://ssrn.com/abstract=2898657

Brown, C. S. (2017). Aligning a performance-based observation rubric to support a teacher performance assessment. *Journal of Interdisciplinary Studies in Education, 5*(2), 11–25. http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=125891944&site=ehost-live

Caliskan, H., & Kasikci, Y. (2010). The application of traditional and alternative assessment and evaluation tools by teachers in social studies. *Procedia - Social and Behavioral Sciences, 2,* 4152–4156. https://doi.org/10.1016/j.sbspro.2010.03.656

Caracelli, J. V., & Cooksy, L. (2013). Incorporating qualitative evidence in systematic reviews: Strategies and Challenges. Mixed methods and credibility of evidence in evaluation. *New Directions for Evaluation, 138*, 97–108. https://doi.org/10.1002/ev.20061

Carrillo, F. J. (2015). Knowledge-based development as a new economic culture. *Journal of Open Innovation: Technology, Market, and Complexity,* 1–17. https://doi.org/10.1186/s40852-015-0017-5

Carson, L., & Kavish, D. (2018). Scaffolding rubrics to improve student writing: Preliminary results of using rubrics in a sociology program to enhance learning and mechanical writing skills. *Societies, 8*(2), 34. https://doi.org/10.3390/soc8020034

Crespo-Martín, B., Figueras-Ferrer, E., Mascarella-Vilageliu, M., Nualart-Torroja, A., & Valle-martí, M. R. J. A. (2016). Proyecto de investigación para la definición e implementación de rúbricas de evaluación de la competencia [Research project for the definition and implementation of competency assessment rubrics]. *REIRE. Revista d'Innovaciói Recerca En Educació Recerca En Educació, 9*(9 (1)), 62–87. https://doi.org/10.1344/reire2016.9.1915

Duckworth, A. L., & Yeager, D. S. (2016). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Research, 44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Duran, R. A., Estay-niculcar, C., Alvarez, H., & Randolph, J. (2015). Exploratory study about the implementation of the competency-based approach in graduate programs from two Panamanian universities. *Journal of Hispanic Higher Education, 1*(18), 1–18. https://doi.org/10.1177/1538192715612913

Ellis, L., Marston, C., Ph, D., Lightfoot, J., Ph, D., Sexton, J., & Ph, D. (2015). Faculty professional development in student learning assessment: *The Assessment Leadership Institute. Research & Practice in Assessment, 10,* 69-80. Retrieved from https://search.proquest.com/docview/1829513087?pq-origsite=gscholar

Elsevier. (2019). *Analyze search results for Rubric along last decade using Scopus Index.*

Escudero Escorza, T. (2016). Evaluation research in the 21st Century: A more and more relevant tool for the educational and social development. *RELIEVE - Revista Electronica de Investigacion y Evaluacion Educativa, 22*(1), 1–20. http://dx.doi.org/10.7203/relieve.22.1.8164

Eshun, E. F., & Osei-Poku, P. (2013). Design students' perspectives on assessment rubric in studio-based learning. *Journal of University Teaching & Learning Practice, 10*(1), 1–13. Retrieved from http://ro.uow.edu.au/jutlp/vol10/iss1/8

Farris, K. B., Demb, A., Janke, K., Kelley, K., & Scott, S. A. (2009). Assessment to transform competency-based curricula. *American Journal of Pharmaceutical Education, 73*(8), 1–14.

Gao, X., & Grisham-Brown, J. (2011). The use of authentic assessment to report accountability data on young children's language, literacy and pre-math competency. *International Education Studies, 4*(2), 41–53. https://doi.org/10.5539/ies.v4n2p41

Gervais, J. (2016). The operational definition of competency-based education. *The Journal of Competency-Based Education, 1*(2), 98–106. https://doi.org/10.1002/cbe2.1011

Goldberg, G. L. (2014). Revising an engineering design rubric: A case study illustrating principles and practices to ensure technical quality of rubrics. *Practical Assessment, Research & Evaluation, 19*(8). Retrieved from http://pareonline.net/getvn.asp?v=19&n=8

Grubb, M. (1981). *Using holistic evaluation* (First). Encino, CA.: Glenco Publishing.

Hancock, D. (2007). Effects of performance assessment on the achievement and motivation of. *Active Learning in Higher Education, 8*(3), 219–231. https://doi.org/10.1177/1469787407081888

Hindriana, A. F., & Setiawati, I. (2018). The development of authentic assessment rubric for assessing undergraduate students' learning and performance. *Indonesian Journal of Learning and Instruction, 1*(1), 21-28. https://doi.org/10.25134/ijli.v1i1.1279

Jin, S. H., Song, K. Il, Shin, D. H., & Shin, S. (2015). A Performance-based evaluation rubric for assessing and enhancing engineering design skills in introductory engineering design courses. *International Journal of Engineering Education, 31*(4), 1007–1020. Retrieved from https://www.researchgate.net/profile/Sung_Hee_Jin/publication/282929432_A_Performance-Based_Evaluation_Rubric_for_Assessing_and_Enhancing_Engineering_Design_Skills_in_Introductory_Engineering_Design_Courses/links/5786009d08aec5c2e4e2c301.pdf

Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment and Evaluation in Higher Education, 39*(7), 840–852. https://doi.org/10.1080/02602938.2013.875117

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2,* 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

78

Kan, A., & Bulut, O. (2014). Crossed random-effect modeling: examining the effects of teacher experience and rubric use in performance assessments. *Eurasian Journal of Educational Research, 14(*57), 1–28. https://doi.org/10.14689/ejer.2014.57.4

Koh, K. H., Tan, C., & Ng, P. T. (2012). Creating thinking schools through authentic assessment: The case in Singapore. *Educational Assessment, Evaluation and Accountability, 24*(2), 135–149. https://doi.org/10.1007/s11092-011-9138-y

Kopp, M. L. (2018). A standardized clinical performance grading rubric: Reliability assessment. *Journal of Nursing Education, 57*(9), 544–548. https://doi.org/10.3928/01484834-20180815-06

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. L. (2014). *Knowing what students know and can do: the current state of student learning outcomes assessment in U. S. Colleges and Universities.* National Institute for Learning Outcomes Assessment. Retrieved from https://learningoutcomesassessment.org/documents/2013%20Survey%20Report%20Final.pdf

Lasater, K. (2006). Clinical judgment development: Using simulation to create an assessment rubric. *Journal of Nursing Education, 46*(11), 496–504. https://doi.org/10.3928/01484834-20071101-04

Latimer, M. E., Bergee, M. J., & Cohen, M. L. (2010). Reliability and perceived pedagogical utility of a weighted music performance assessment rubric. *Journal of Research in Music Education, 58(*2), 168–183. https://doi.org/10.1177/0022429410369836

Lee, V. R., & Fields, D. A. (2017). A rubric for describing competences in the areas of circuitry, computation, and crafting after a course using e-textiles. *International Journal of Information and Learning Technology, 34*(5), 372–384. https://doi.org/10.1108/IJILT-06-2017-0048

Mertler, C. A. (2001). Designing score rubrics for your classroom. *Practical Assessment, Research and Evaluation, 7(*25), 1–8. https://scholarworks.umass.edu/pare/vol7/iss1/25

Meyer-Adams, N.; Potts, M. K.; Koob, J.; Dorsey, C. J.; Rosales, A. M. (2011). How to tackle the shift of the educational assessment from learning outcomes to competencies: one program's transition. *Journal of Social Work Education, 47*(3), 489–508. https://doi.org/10.5175/JSWE.2011.201000017

O'Brien, C. E., Franks, A. M., & Stowe, C. D. (2008). Multiple rubric-based assessments of student case presentations. *American Journal of Pharmaceutical Education, 72*(3), 1–7. https://doi.org/10.5688/aj720358

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9,* 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448. https://doi.org/10.1080/02602930902862859

Rodriguez, I., & Gallardo, K. (2017). Redesigning an educational technology course under a competency-based performance assessment model. *Pedagogika, 127*(3), 186–204. http://dx.doi.org/10.15823/p.2017.48

Roy, G. (2016). *Competency-based Assessment in Primary Education in Bangladesh- A Review.* Retrieved from https://ssrn.com/abstract=2899109 or http://dx.doi.org/10.2139/ssrn.2899109

Schoepp, K., Danaher, M., & Kranov, A. A. (2018). An effective rubric norming process. *Practical Assessment, Research & Evaluation, 23*(11), 1-12. https://scholarworks.umass.edu/pare/vol23/iss1/11

Slaus, I. (2013). Entering global knowledge society: Role of education. Donald School. *Journal of Ultrasound in Obstetrics and Gynecology, 7,* 239–247. https://doi.org/10.5005/jp-journals-10009-1289

Smit, R., Bachmann, P., Blum, V., Birri, T., & Hess, K. (2017). Effects of a rubric for mathematical reasoning on teaching and learning in primary school. *Instructional Science, 45*(5), 603–622. https://doi.org/10.1007/s11251-017-9416-2

Tigelaar, D., & Janssen, F. (2012). Dilemmas in designing rubrics for the assessment of student teachers in teacher education: Stories of a design and evaluation process in the context of biology teaching. *Educación Química, 23*(2), 179–187. https://doi.org/10.1016/S0187-893X(17)30106-4

Velasco-Martínez, L. C., & Hurtado, J. C. T. (2018). The use of rubrics in higher education and competences evaluation. *Profesorado, 22*(3), 183–208. https://doi.org/10.30827/profesorado.v22i3.7998

Verano-Tacoronte, D., González-Betancor, S. M., Bolívar-Cruz, A., Fernández-Monroy, M., & Galván-Sánchez, I. (2016). Valoración de la competencia de comunicación oral de estudiantes universitarios a través de una rúbrica fiable y válida [Assessment of the oral communication competence of university students through a reliable and valid rubric]. *Revista Brasileira de Educação, 21*(64), 39–60. https://doi.org/10.1590/S1413-24782016216403

Voronov, M., Chasovskykh, V., Popov, V., Anyanova, E., & Krainova, T. (2015). Effectiveness research of the new learning elements initiated by the change to competency-based education model in Russia. *Procedia - Social and Behavioral Sciences, 214,* 758–768. https://doi.org/10.1016/j.sbspro.2015.11.712

Yune, S. J., Lee, S. Y., Im, S. J., Kam, B. S., & Baek, S. Y. (2018). Holistic rubric vs. analytic rubric for measuring clinical performance levels in medical students. *BMC Medical Education, 18*(1), 1–6. https://doi.org/10.1186/s12909-018-1228-9

Katherina Gallardo   Dr., Director of Educational Innovation Ph.D. Program, School of Humanities and Education, Monterrey Institute of Technology and Higher Education, Av. Eugenio Garza Sada 2501 Sur. Colonia Tecnologico, Mexico.
E-mail: katherina.gallardo@tec.mx
Website: https://itesm.academia.edu/KatherinaGallardo
ORCID: https://orcid.org/0000-0001-8343-9518