THE ROLE OF RUBRICS IN LEARNING AND IMPLEMENTATION OF AUTHENTIC ASSESSMENT: A LITERATURE REVIEW

Clara Nkhoma	RMIT University Vietnam, Ho Chi Minh City, Vietnam	clara.nkhoma@rmit.edu.vn
Mathews Nkhoma*	RMIT University Vietnam, Ho Chi Minh City, Vietnam	mathews.nkhoma@rmit.edu.vn
Susan Thomas	RMIT University Vietnam, Ho Chi Minh City, Vietnam	susan.thomas@rmit.edu.vn
Nha Quoc Le	RMIT University Vietnam, Ho Chi Minh City, Vietnam	quocnhale@gmail.com

ABSTRACT

Aim/Purpose	This study will review the existing literature on the advantages and challenges associated with rubric design and implementation. The role of rubric as an authentic assessment instrument will also be discussed.
Background	This study provides an overall understanding of 'rubric design, the benefits and challenges of using rubrics, which will be useful for both practitioners and researchers alike.
Methodology	A comprehensive literature review was carried out on rubric, educational assessment, authentic assessment and other related topics.
Contribution	Different types of rubrics and essential elements to create a complete rubric for classroom effectiveness are reviewed from literature to aid researchers, students and teachers who are new to using and designing rubrics. For experienced rubric designers and users, this will be an opportunity for them to get reassurance from the literature regarding good practices of rubric usage. This project will also be of use to researchers working on rubrics.
Findings	A rubric is not only an assessment tool useful for students in high-stakes exam but also an educational instrument supporting learners to select appropriate learning approaches, assisting teachers to design effective instruction strategies, and improve reliability and validity of assessment.

	Novice learners should begin with generic rubrics due to their simplicity. Meanwhile, a task-specific rubric is more useful to improve reliability and validity of large-scale assessment. Holistic rubrics are appropriate for assessment of learning and analytic rubrics are almost indispensable in student-centred classroom and assessment for learning.
Recommendations for Practitioners	A rubric as an authentic assessment instrument is useful to enhance the reliability of authentic assessment. Moreover, other empirical results indicate that rubrics play an importance role in authentic assessment regardless of levels or disciplines.
Recommendations for Researchers	Those carrying out research on rubrics, rubric design and authentic assessment will find this paper useful as a point of reference to inform their research.
Impact on Society	The findings apply to both learners and instructors in terms of analyzing best practices when using rubrics. The paper highlights that there are three main factors that determine the effectiveness of a rubric in improving students' performance: namely, the users' perception, the design, and the purpose of using rubrics. Rubric designing variables should also be optimized based on reliable data and information about the target educational context. Armed with this information, instructors will be in a better position to optimise the learning experience of their students.
Future Research	Systemic literature reviews with data analysis from both qualitative and quantitative findings should be carried out in the future to identify current trends and the role of rubrics in learning.
Keywords	rubric, authentic assessment, rubric design, assessment method, scoring guide

INTRODUCTION

WHAT IS A RUBRIC?

In the early of 1400s, the term "rubric" was initially used in English to refer to the colour red. Rubric also appeared in antique writings, catholic archives and law documents where it generally meant "a noted sentence of any book marked with red letters" (Cooper & Gargan, 2011). However, that was only the root of an interesting story of rubrics. Nowadays, 'rubric' is widely known and studied as an educational term. There are many studies that have been done on rubric design, benefits and disadvantages associated with rubric usage. This implies that in the field of education, a rubric is not merely a book marked with red letters.

There is a wide variety of definitions of a rubric in education. To some extent, rubrics are scoring schemes (Popham, 1997) that help and guide people in judging a variety of constructs such as quality of students' work, academic performance as well as educational resources (Moskal & Leydens, 2000; Porcello & Hsi, 2013; Rezaei & Lovorn, 2010). According to Andrade (2000) and Arter and Chappuis (2006), a rubric lists the scoring criteria and portrays all levels of quality. Jönsson and Panadero (2016) define rubrics as "assessment instruments designed to assist in identifying and evaluating qualitative differences in student performance". In other words, whatever rubrics are called (rules, guides, criteria or descriptions, etc.), they are deliberately designed to clarify what is needed to reach different levels of quality.

A rubric is supposed to be specific by the elements of its structure, which nail down the essential information of the rubric. Herman (1992) highlighted four basic elements in rubric design whereas

Dawson (2017) proposed fourteen design elements. In terms of format, information in a rubric can be presented in either table format or info-graphic, with emojis or samples and with quality words, quantity numbers or even detailed explanations. These varieties will probably bring flexibility to educators as well as rubric designers. However, in the absence of a consistent definition, previous research on rubrics has usually caused divergence of interpretation (Dawson, 2017). Therefore, this study employs the definition of Allen & Tanner (2006) that rubrics or assessment rubrics are "a type of matrix providing scaled levels of achievement or understanding with detailed descriptions for a set of criteria or dimensions of quality and scoring strategy for a given type of performance".

BENEFITS OF USING RUBRICS

In recent years, mainstream research has focused more on the use of rubrics in education to investigate how and to what extent they can benefit students in learning and achieving better results as well as how they assist teachers and educators in making accurate assessment about their students' performance. Rubrics are basically supposed to help students understand teachers' expectation and make the scoring more consistent by clarifying fuzzy goals (Andrade & Du, 2005; Panadero & Jönsson, 2013; Sundeen, 2014; Wolf & Stevens, 2007).

Clear evaluation of criteria and explicit quality definitions for those criteria at given levels of achievement are key components of rubrics that promote students learning. Being well informed about what is important enables students to design their own learning plans more effectively. Sitting in the same classes or even the same courses does not mean students hold identical aptitudes, backgrounds and career goals. As a component of student-centred approach of education, rubrics provide students with clear direction and targets thereby helping them map out an optimal learning strategy (Alonso-Tapia & Panadero, 2010; Andrade & Du, 2005; Li & Lindsey, 2015; Schmoker, 2006; Wiggins, 1998). In other words, when students are under constrains of time, cost, capacity and learning-orientations, a good understanding of criteria of evaluation allow them to allocate their effort in different subjects adequately. Acting as a self-assessment tool, rubrics simplifies learners' self-monitoring; allow them to get peer-feedbacks immediately during collaborative learning and doing assignments (Bolton, 2006; Jönsson & Panadero, 2017; Linn et al., 1991). In many relevant studies, immediate feedbacks are considered especially useful to low-ability students who are struggling with difficult tasks to achieve higher performance (Attali & Van der Kleij, 2017; Bangert-Drowns et al., 1991). They are also supposed to enhance cognitive and metacognitive performance, calibration accuracy, self-efficacy as well as self-regulated learning skills, engagement, motivation, and learning outcomes (Corbett & Anderson, 2001; El Saadawi et al., 2008; Feingold et al., 2008; Muis et al., 2015; Nietfeld et al., 2006; Parmelee et al., 2012; Rawekar et al., 2013; Schmid et al., 2008; Schunk, 2003; Stone, 2000; Winne & Hadwin, 1998; Zimmerman, 2000). Timmerman et al. (2011) carried out a survey on 1026 observations and received the result that a great proportion of students considered peer feedback based on rubric criteria improved the quality of their papers. Rubrics are not only useful in assessment for learning but also in high-stakes tests. Including a transparent scoring strategy, they help students better prepare for exams, diminish anxiety and reduce arguments about the final results (Andrade & Du, 2005; Wolf & Stevens, 2007). Students may use the information in rubrics to check the accuracy of assessments and therefore, the confidence and fairness of their grades are improved.

Through the scope of teachers and educators, rubrics can also benefit in many aspects. Firstly, rubrics can be used for a wide variety of learning contexts such as virtual learning, essays, lab work, presentation, exhibits and performances, portfolios of student work, artwork and even internships (Karkehabadi, 2013). Menéndez-Varela and Gregori-Giralt (2016) discover that by involving rubrics in distance learning courses, students' achievement and satisfaction is better enhanced. Therefore, they suggest considering rubrics not only as a scoring tool but also as a first-order teaching resource. Secondly, with the help of rubrics, the judgment process probably becomes more accurate, unbiased and consistent (Wolf & Stevens, 2007). When students are enjoying their trips to the beach, throw a house party or doing barbeque by the swimming pool after finishing exams, their teachers may be

buried under tons of answer sheets. Subjective opinion is a potential factor diminishing an assessment's accuracy. Sometime, teachers even mobilize all their teaching assistants to mark exam papers. Without specific performance criteria provided in rubrics, it is impossible to control the validity and reliability in scoring process (Wolf & Stevens, 2007). Furthermore, the rubric appears to be an efficient reference for educators to design suitable learning environment as well as assessment methods (Arter & McTigue, 2001). It supports instructors to detect weaknesses or misalignments in curricular (Halonen et al. 2003). A well-designed learning environment is a key to enhance teachers' confidence and students' motivation, which in turn has a positive impact on academic performance (Coe et al., 2006; Hillman et al., 2008; Hillman et al., 2009). That is why Song (2006) concluded in his study that rubrics can teach as well as evaluate. After decades with hundreds of empirical studies to investigate how rubrics can benefit learning and teaching activities, there are inevitably some results reporting negative effect due to preventing students from learning further than what is mentioned in rubrics (Wiggins, 1998). However, providing a well-designed rubric with explicit learning targets even before the courses start is widely considered important for both learner and educator to avoid time-consuming activities, inconsistency and biasness in learning and assessment process. It seems to be the most terrible disaster that no student at any grade expects to encounter is not boring classes, tough assignments, and stressful examinations but the moment they realise what they learn is irrelevant to what appears on their question papers.

TYPES OF RUBRICS

The pedagogical terminology "rubric", in general, refers to "a scoring guide used to evaluate the quality of students' constructed responses" (Popham, 1997). They are typically presented in table format and consist of four basic elements (Herman et al., 1993); namely, scoring criteria, a rating scale, definitions of each criterion and descriptions for specified performance levels (see Table 1). However, there are several types of rubrics specified by their composition and contend to fulfil the educational objective of the courses. In terms of particular object of assessment and scoring strategy, rubrics are principally categorized into generic-rubric, task-specific rubric, holistic and analytic rubric (Dawson, 2017).

As mentioned by Tierney and Simon (2004), rubrics are either "generic" or "specific" according to their level of specificity. Generic rubrics are shaped with performance criteria designed to reflect broad learning targets (Tierney and Simon, 2004). For example, a problem-solving rubric is useful in dealing with assignments on Math, Physics, Economics, and so forth; a reading rubric can be applied not only for literature courses. In fact, students often face similar learning processes when approaching different domains. With criteria that are general across different context, educators can save considerable time on creating rubrics for each specific task (see Table 2), students may easily capture the essential principles across disciplines as well (Bargainnier, 2003). At the opposite pole, task-specific rubrics are equipped with criteria and descriptions that represent specific features of a performance. Therefore, it is applicable for only one particular task, football dribbling for instance (Bargainnier, 2003). That narrow space of application is the trade-off for more reliable assessment and concrete descriptions to guide interpretation (Tierney and Simon, 2004). It is worth noting that there is an open room for teachers who have the intention to harmonize the drawbacks of generic and specific rubrics without sacrificing their advantages. That hybrid rubric may contain generic language and some criteria that are task-specific (Center for Advance Research Language Acquisition, 2018).

Table 1. Sample of generic rubric
(Generic rubric for practice)

Grading Criteria	Poor (1)	Below Average (2)	Average (3)	Good (4)	Excellent (5)
Organization	The essay is unclear with no organization.	The main points of the essay are ambiguous.	Writing has minimal organization and a basic thesis statement.	Writing follows a logical organization, but sometimes drifts from the thesis.	Writing is clear, logical, and very organized around a developed thesis.
Evidence	The essay does not attempt to use evidence to support thesis.	The evidence provided does not support thesis.	The use of evidence is minimal, but it does support thesis.	There is evidence to support almost every point.	Every point is clearly supported by strong evidence.
Analysis	The essay does not attempt to explain how the evidence relates to thesis.	The analysis of the evidence has no relation to the thesis.	The analysis of the evidence stretches its meaning to support thesis.	The analysis explains how the evidence supports the thesis in most cases.	The analysis shows a strong relationship between the evidence and the thesis.

Source: http://gsi.berkeley.edu/media/generic-essay-rubric.pdf
(Graduate Student Instructor – Berkeley Graduate Division)

Table 2. Sample of task-specific rubric
(Dribbling rubric)

Point Level	Description of Dribbling Ability
0 points	Cannot perform the skill.
1 point	Can control a ball through four cones three feet apart in more than ten seconds.
2 points	Can control a ball through four cones three feet apart in ten seconds.
3 points	Can control a ball through four cones three feet apart in seven seconds.
4 points	Can control a ball through four cones three feet apart in five seconds

Source: http://www.cwu.edu/~gossge/curriculum/rubric/rubrics
(Central Washington University)

Rubrics are also categorized as "holistic" and "analytic" based on the procedures of grading and marking. When component parts of a task need to be judged separately, analytic rubric is an ideal option. The rubric is often presented as a matrix with performance criteria of a single task are listed in the leftmost column, the top row lists the levels of proficiency and the descriptions are stored in the cells. The users obtain the overall score by summing the results from all criteria (Moskal & Leydens, 2000; Nitko, 2001). This composition allows the assessors to put different weights in different fea-

tures and, therefore, students are clearly aware of which modules of the task they should place emphasis on. The learners can also use analytic rubrics as a flow chart to track the progress they have made over time (Moskal & Leydens, 2000) and self-assess their capacity (Bargainnier, 2003). The multidimensional matrix gives such specific feedback that it can be a reliable source to develop students' profiles of strengths and weakness (Mertler, 2001). Creating and applying such a detailed rubric is obviously time-consuming (Mertler, 2001) and more possibilities for raters to disagree (McNamara, 1996). In contrast, a holistic rubric requires the teacher to score the overall process or product, without judging the component parts separately (Nitko, 2001). The rubric looks like a uni-dimension matrix or a text-based list of levels (see Table 3) with each band on the scale describing performance on multiple traits (CARLA, 2018). This holistic scoring strategy enables the measurers to focus on the overall quality, proficiency, or understanding of the specific content and skills (Mertler, 2001). Thus, it is supposed to be more appropriate with open-answered tasks, when the "whole is greater than the sum of its parts" (Nitko, 2001). It is quick to use and easy to construct a holistic rubric but the lack of specific details and limitation in providing feedback for improvement makes it not useful for formative assessments (CARLA, 2018; Mertler, 2001).

Table 3. Sample of holistic rubric
(Articulating thoughts through written communication)

Score	Description
4	*Above Average:* The audience is able to easily identify the focus of the work and is engaged by its clear focus and relevant details. Information is presented logically and naturally. Mechanical errors or misspelled words do not distract the reader.
3	*Sufficient:* The audience is easily able to identify the focus of the student work which is supported by relevant ideas and supporting details. Information is presented in a logical manner that is easily followed. Minimal interruption to the work due to misspellings and/or mechanical errors.
2	*Developing:* The audience can identify the central purpose of the student work without little difficulty and supporting ideas are present and clear. The information is presented in an orderly fashion that can be followed with little difficulty. There are some misspellings and/or mechanical errors, but they do not seriously distract from the work.
1	*Needs Improvement:* The audience cannot clearly or easily identify the central ideas or purpose of the student work. Information is presented in a disorganized fashion causing the audience to have difficulty following the author's ideas. There are many misspellings and/or mechanical errors that negatively affect the audience's ability to read the work.

Source: https://curry.virginia.edu/uploads/resourceLibrary/Info_on_Rubrics.docx
(Curry School of Education and Human Development – University of Virginia)

Rubrics not only guide the scoring process but also provide qualitative feedback. The single-point rubric is a typical one. With the pedagogical philosophy that descriptive feedback is indeed more valuable than a number or a label of proficiency for setting improvement plan (Fluckiger, 2010), a single-point rubric consists of only guidance on and descriptions of successful work (Hashem, 2017). In more detail, only the descriptions of proficient level of criteria are informed and placed in the middle column. The leftmost and rightmost columns are left blank for notes on what needs improvement and how the work exceeds expectations respectively (Gonzalez 2015). This three-column format creates more time and space for student to identify and note their own strengths and weaknesses in work by minimizing the use of text. That unique feedback prevents them from ranking and comparing themselves with one another and helps take student's attention off the grade (Hashem, 2017). Moreover, it saves teachers' time on attempting to predict and prescribe all level of mastery as noted

by (Fluckiger, 2010) that "students may surprise us if we leave quality open-ended". Developmental rubric is another type of rubric in which the central purpose is not scoring an end performance. Intuitively, a developmental rubric is a subset of analytic rubrics with multiple traits of evaluation (see Table 4). Its primary goal is to evaluate how students are developing their skills and proficiency. Based on the relevant developmental theory, the rubric's design is characterized and distinguished by the set of criteria. Hence, it is quite challenging to design an applicable developmental rubric.

Table 4. Sample of analytic rubric

	Value 3	*Value 2*	*Value 1*	*Score*
Describe	Provides a complete and accurate description of the key subject matter and elements seen in the artwork.	Provides a partial but mostly accurate description of the subject matter and/or elements seen in the artwork; some key components overlooked.	Provides an incomplete, unclear, or inaccurate description of subject matter and/or elements seen in the artwork; many key components overlooked.	
Analyze	Accurately relates how the structures of art function together to make a complete composition.	Relates with limited proficiency how the structures of art function together to make a complete composition; overlooks some important components.	Has trouble relating how the structures of art function together to make a complete composition.	
Interpret	Suggests a logical and/or symbolic meaning expressed in a work of art; supports idea with multiple points of visual evidence found in the piece.	Suggests a literal meaning expressed in a work of art; supports idea with limited points of visual evidence found in the piece.	Finds it difficult to interpret the meaning of the work; guesses meaning without visual support.	
Evaluate	Uses multiple criteria to judge the quality of a finished work of art; avoids personal opinion.	Uses a limited range of criteria to judge the quality of a work of art; personal opinion shown.	Uses personal opinion to judge the quality of a finished work of art.	
Technical	Finished paper follows rules of grammar and essay writing; is in publishable form.	Finished paper contains minor flaws in grammar and essay writing; needs editing.	Finished paper has numerous flaws in grammar and does not follow conventions of essay writing; needs rewriting.	
				Total score

Source: https://www.academia.edu/30309500/Analytic_Rubric_Sample_From_Rubrics_The_Heart_of_Assessment

GENERIC RUBRIC: A SELF-ASSESSMENT TOOL

The rubric is an important component of authentic assessment (Menéndez-Varela & Gregori-Giralt, 2016). It provides transparent grading criteria, useful feedback and clarifies the educational goal of the course, which in turn makes the learning process more meaningful and reinforces the confidence

of assessment. However, there is no such thing as one-size-fits-all when it comes to rubric application. According to the learning objective, instruction strategy and assessment method, educators choose the most appropriate type of rubric to facilitate their educational activities. In practice, essential benefits of generic rubrics are better understood when it is used by novice learners who are struggling with fundamental skills (Brookhart, 2013; CARLA, 2018).

Teachers may face difficulties in designing effective instruction schemes if they lack understanding the learning behaviours of students. The first step is always the hardest. Therefore, learning behaviours of novice learners is a topic that has attracted several studies (Chi et al., 1988; Swanson et al., 1990). In general, they are supposed to be less aware of when they need to check for errors, why they fail to comprehend, and how they need to redirect their efforts (A. L. Brown & DeLoache, 1978). Moreover, allocating their time, monitoring their own thinking and problem solving are also weaknesses of novice learners (Bransford et al., 1989). The generic rubric is born to assist learners overcome those hurdles in principle. The rubric provides general information about essential ingredients of the 'skill to be measured' (Popham, 1997; Oakleaf, 2006) and contains descriptions that help students focus on what their learning target is supposed to be (Learnalberta.ca, 2018). Students, thereby, can employ the rubric as a lighthouse to keep their learning on track. Time and effort allocation now appear to be simplified when students are fully aware of where they need to place emphasis on. Besides, the key feature that distinguishes generic rubrics with task-specifics ones is the ability to be applied in many different tasks, as long as they belong to the same domain because no specific features of a particular task is included in a generic rubric. Theories, concepts, task directions or answers to a problem will not be found in generic rubrics (Brookhart, 2013). In fact, a well-built generic rubric aims at encouraging students develop their skills and abilities across different contexts (Learnalberta.ca, 2018). Thus, novice leaners will not be overloaded with too much information (Moreno, 2004), which is usually associated with a loss of control over the situation, and sometimes with feelings of being overwhelmed (Bawden & Robinson, 2009). General descriptions of characteristics of strong performance help students build up a concept of what it means to perform a skill well. Brookhart (2013) writes: "when the rubrics are the same each time a student does the same kind of work, the student will learn general qualities of good essay writing, problem solving, and so on". Bearing in mind this concept when dealing with different assignments, students can explore their own paths to success as well as self-evaluate the knowledge and skills (e.g., problem-solving) they are developing over time (Brookhart, 2013; Tierney & Simon, 2004), which is also mentioned by Petkov and Petkova (2006) as long-term assessment and comparability function of generic scoring rubric.

However, there are some challenges that need to be considered when a generic rubric is applied. Balan (2012) recorded positive effects on students' learning when a generic rubric is used to support mathematical problem solving. Nevertheless, it took the students a couple of weeks to comprehend and use the rubric for self-assessment and peer-assessment purposes. One possible reason is that novice learners are unlikely to use self-tests and self-questioning as sources of feedback to correct misconceptions and to redirect the use of learning strategies (Bransford et al., 1982; A. L. Brown et al., 1983; Rafoth et al., 1993). Therefore, Brookhart (2013) concludes that practice is required to better apply the rubric. Moreover, Orsmond et al. (1996) and Jönsson and Svingby (2007) argue that students' lesser developed sense of how to interpret criteria may cause differences between instructor and student judgments. In the same vein, Li and Lindsey (2015), in their effort to explore the variations between students' and teachers' readings of rubrics, realized that discrepancies in interpretation and application of the rubric in essay evaluation are potential limitations, especially in first-year writing programs. For this reason, Wiggins (1998) and Busching (1998) suggest that the variability of students and teachers' interpretation can be reduced significantly when generic terms are clarified with "anchors", task-specific exemplars or indicators. In order to make both the student and teacher confident that they are developing the required skills and understandings, specific features of understanding included in a generic rubric should also correlate to learning outcomes in the program of studies (Learnalberta.ca, 2018).

TASK-SPECIFIC RUBRIC: A SCORING GUIDE TO IMPROVE ASSESSMENTS' RELIABILITY

It is quite plausible if task-specific and generic rubrics are compared with two tails of a coin. A generic rubric contains general criteria and is reusable across tasks whereas task-specific rubric reflects specific features of the elicited performance (Brookhart, 2013). While the former helps students capture fundamental and general principles, the latter details elements required to complete a particular task successfully (Bargainnier, 2003). However, both rubrics can well complement each other to fulfil the need of students from elementary to a more specialized level of knowledge. When a task-specific rubric is involved, students seem to score higher, reach further achievement (Howell, 2011; Howell, 2014) and obviously, the assessment's reliability is also improved (Brookhart, 2013). Nevertheless, only those responsible for grading in large-scale and authentic assessment fully appreciate what task-specific rubrics can do.

According to Gulikers et al. (2004), learning, instruction and assessment need to be closely aligned in order to attain educational goals. This implies traditional knowledge transmission is only compatible when knowledge acquisition is the primary goal of learning. Modern education philosophy focus on analytical thinking and competence development, which "empower learners for the jobs and skills of the future" (Deakin University, 2017, p. 9), hence, requires more practical instructions and an authentic assessment approach. Gulikers et al. (2004) argued that authentic assessment requires students to use "the same competencies, or combinations of knowledge, skills, and attitudes, that they need to apply in the criterion situation in professional life" and, therefore, impacts on the quality and depth of learning achieved by the student (Dochy & McDowell 1997; Wiggins 1993) as well as enhances employability (Wu et al., 2015). Although it is quite appropriate for modern learning contexts with countless advantages, one of the most serious downsides of authentic assessment is subjectiveness (Ewing, 1998; Honebein et al., 1993; H. M. Huang, 2002; Petraglia, 1998), which means that the perception of student and teacher or assessment developers about authentication is not homogeneous (Gulikers et al., 2004). As a result, this heterogeneity in perception may cause problems related to reliability and validity of scoring. Reliability implies how an assessment is independent of who does the scoring (Jönsson & Svingby, 2007) and validity, according to Jönsson and Svingby (2007) in this context, answers the question "Does the assessment measure what it was intended to measure?". When it comes to the problems of reliability and validity, a task-specific rubric is considered as a feasible solution. A rubric with clear learning targets and confirms criteria with explicit descriptions of specific tasks will assure students and assessors hold a similar view of authentic assessment. In fact, the lack of consistency of an individual marker, which is supposed to be a "major threat of variation in judgments" (G. Brown et al., 1997), may not be a considerable concern when those marking are supported by a rubric. In the same line, Arter and McTighe (2001) declared that more specific scoring rubrics can transform a subjective evaluation process into a clear, consistent, and verifiable procedure.

It is obvious that assessors may draw different conclusions about the same performance because of differences in experience and disagreement on scoring routines (Jönsson & Svingby, 2007). Thus, a task-specific rubric is especially useful in high-stake assessments and especially large-scale assessments, when student progress is measured at the local, state or national level. The rubric makes scoring easier for teachers, require less time and a minimum of practice to achieve inter-rater reliability (Brookhart, 2013; Pindiprolu et al., 2005). On the other hand, designing a good quality task-specific rubric is difficult and requires great investment in terms of time and effort. For this reason, applying task-specific rubric in large-scale assessments, which include hundreds or thousands of students, might help solve the problem of cost-benefit optimization. Moreover, involving experts in task-specific rubric development is a recommended solution to overcome its quality requirement as well as improve the rubric content validity (Taggart et al., 2001). Finally, when the criteria of a task are more specifically defined, assessors may find it awkward when they encounter open-ended tasks. Pindiprolu et al. (2005, p. 2) foresee this issue and suggest: "it is important to develop a generic scoring rubric that delineates objective scoring criteria that reflects current best practices in the field and then to use the generic scoring rubric guidelines to develop task specific scoring rubrics to help assess a

student's performance on a case". Generic and task-specific rubrics are again deemed to be two sides of coin.

HOLISTIC RUBRIC: WHEN THE WHOLE IS GREATER THAN THE SUM OF ITS PARTS

One of the basic principles of designing rubrics is that after rubric developers make up their mind on what level of specificity is appropriate, the procedure of scoring or marking should be put under consideration (Bargainnier, 2003). The question is thus, in what context are analytic and holistic rubrics useful for? According to Young (2013), different degrees of judgment complexity might be more useful in different stages of learning. He argues a complex rubric appears to be more useful for learning, but a holistic rating scale was found easier to use once the learning had occurred. The statement indirectly implies that assessment of learning is the target of holistic rubrics.

As emphasized by Gibbs (1992), assessment and student perceptions of the assessment are two major factors determining student learning. Under the viewpoint of Nasab (2015), assessment is "a method of following students' progress through active participation of the learner himself" and can be categorised in three kinds: namely, assessment for learning, assessment as learning, and assessment of learning. The author defines assessment of learning as the process of ascertaining what the students know in relation to curriculum outcomes. The definition is like that of summative assessment, whose prior purpose is to spot what students know at the end of the term (McAlpine, 2002). From this point, it is not difficult to recognize that summative assessment or assessment of learning is the rich land for holistic rubric. Technically, the rubric requires its users to do the scoring by evaluating all criteria simultaneously (Brookhart, 2013). With band scores of multiple criteria, holistic rubric leads the measurers to focus on the overall impression of students' performance (Mertler, 2001). This attitude of assessment also allows holistic rubrics to be more compatible for performance and authentic assessment. In traditional classrooms, the testing culture is supposed to consist of decontextualized, psychometrically designed items in a choice-response format to test for knowledge and low-level cognitive skill acquisition (Gulikers et al., 2004). Meanwhile, authentic assessment in present-day education puts a special emphasis on realistic value and therefore, it is characterized by open-ended tasks. A holistic rubric does not judge students by how the learners accomplish the task step by step and what they cannot do but instead evaluate the impact of their product or performance.

Moreover, Brookhart (2013) argues that with the presence of holistic rubric in an assessment, reliability and validity are more likely to be guaranteed because the rubric minimizes the number of decisions assessors must make. However, this statement needs to be further discussed before a conclusion is drawn. Holistic rubric lets the assessors evaluate all criteria simultaneously. On the one hand, it is not possible to put different weights on the criteria. On the other hand, that also means the scorers can place emphasis on those criteria subjectively, which in turn may cause both inter-rater and intra-rater reliability issues. Another drawback is that students can only get limited feedback with the reference of holistic rubrics. In fact, they receive nothing but a final score or grade with general descriptions for that level. Due to these shortcomings, most holistic rubrics in practice is constructed for a specific simple task. This strategy brings many benefits in terms of reliability improvement. Empirical results of this design are positive as well. When Wallace et al. (2011) observed 21 undergraduate students who were provided with Astronomy–Cosmology holistic task-specific rubrics found evidence of an increase in reliability of measurement. In addition, 80 students in the sample of Howell (2011) got higher scores and better performance when a holistic task-specific rubric was employed.

ANALYTIC RUBRIC: A CRUCIAL PIECE TO OPERATE A STUDENT-CENTERED CLASSROOM

The Irish poet and playwright, the winner of Nobel Prize for literature William Butler Yeats said, "Education is not the filling of a pail but the lighting of a fire." Learning should not be a boring and stressful process for students. It sounds quite different to the atmosphere in teacher-centered classrooms in which students work independently and not just rely on what teachers are saying (Slunt &

Giancario, 2004). In fact, active collaboration is strongly encouraged in classrooms operating with student-centered instruction (Brush & Saye 2000). Collaborative learning, as described by innovative pedagogical educators, may lead to many advantages with recorded positive influence in college environments, in classroom and student satisfaction (Barkley et al., 2014; Laal & Ghodsi, 2012). In a learning environment where both students and instructors share the focus, where interaction and students' interest are priorities and where meaningful feedback is essential, analytic rubric is indeed an irreplaceable learning tool.

In the past, teachers used to hold a central position in the classroom and even in the learning process of their students. They played a role as a didactic information giver and controlled the academic content. This asymmetry of power between learners and teachers might keep the classroom quiet and ordered but as a trade-off, students are not allowed to express themselves, ask questions, and direct their own learning. The student-centered approach, in contrast, enables learners to choose not only what to study but also how they will learn, why that topic might be of interest and how they will assess their own learning (Hannafin & Hannafin., 2010; C. R. Rogers, 1983). With regard to the ultimate goal of learning, student-centered learning, also known as learner-centered education, aims to develop learner autonomy and independence (Jones, 2007). Therefore, students are strongly required to actively communicate and collaborate with their peers as well as their teachers. When students are placed in an open-collaborative environment, both participation engagement and group interaction are enhanced (Fredricks et al., 2004; Vaughan, 2014; Visschers-Pleijers et al., 2006), which increases their individual learning performance (D. W. Johnson & Johnson, 2009; Krause & Coates, 2008). To construct a lifelong learning and independent problem-solving habit, students are also enabled to create their own activities and select their own authentic materials (Corley, 2008; Young, 2007). The philosophy of student-centered education also needs to be reflected in the method of instruction. By placing more emphasis on students' activities, the instructors in student-centered classrooms act as facilitators who encourage students to do more discovery learning and to learn from each other rather than trying to "fix" them (McCombs & Whistler, 1997; Weimer, 2002). The curriculums are therefore constructed with authentic, real-life tasks that motivate learner involvement and participation (Weimer, 2002). Finally, moving from teacher-centered learning toward learner-centered education transforms the function of assessment (Crumly & Dietz, 2014). Student-centered learning places more emphasis on formative assessment (Crumly & Dietz, 2014) and consider the involvement of students in the evaluation of their learning as an important feature to the success of the student-centered approach. Students' performance has been demonstrated to be positively improved when the focus is shifted from teachers to learners (Wright, 2011)

However, learners and educators should not be overwhelmed by the benefits and ignore challenges they may face. The modification process of learning, instruction and assessment mindsets is always in need of assistant tools to assure compatibility. An analytic rubric is certainly one of the most interesting names to be considered. In principle, the analytic rubric provides detailed descriptions of every single criterion at all levels. Students with the rubric in hand obviously have a clearer vision of what they need to achieve and actively map out their own paths to reach the targets even before the courses begin. By dividing a product or performance into essential dimensions, the analytic rubric allows the designers to assign different weights to different features (Tedick, 2002). As a result, students' learning plans and decisions will relate more to educational objectives (Truemper, 2004).

Different from the holistic rubric, transparent framework and scoring guides, which evaluate each criterion separately, are beneficial to both the faculty and student. On the one hand, teachers can employ the rubric to draw summative assessment at the end of the courses. On the other hand, this inductive scoring method provides detailed feedback about the learners' performance. Shipman et al. (2012, p. 1) writes: "If students are not given feedback that is consistent or meaningful, then they may feel their grades are not valid, thus sending mixed messages to the student. If expectations are not clear, the unacceptable performance may transfer into the work environment." Similarly, Nasab

(2015, p. 3) argued, "It's a diagnostic approach which aims at providing meaningful feedback to facilitate students' learning and improve teachers' teaching". Assessment that not only reports overall score but also provides detailed feedback is defined as assessment for learning (Nasab, 2015) or also known as formative assessment (Black and Wiliam, 1998). Assessment for learning offers many opportunities for students to develop their own skills by making evaluations about their own performances (Race et al., 2004). In a learning setting where teachers do not always deliver instruction to all students at once, understanding key facts may become problematic. Thus, self-assessment is even more important in the student-centered environment by equalizing understanding and avoiding missing important instructions for all students (Stevens and Levi, 2005).

With a set of scores on all dimensions, teachers are enabled to construct individual profiles of students' strength and weakness (Mertler, 2001). In the learner-centered mindset, learning is a constructive process that is relevant and meaningful to the learner and connected to the learner's prior knowledge and experience and each learner is distinct and unique (Corley, 2008). Hence, this kind of profile plays a crucial role in a student-centered environment where instructors are required to be well aware of and respect learners' backgrounds, interests, abilities, and experiences (McCombs & Whistler, 1997). Furthermore, these profiles can be employed as reliable data resources to students' achievement forecasting, behaviour analysis as well as quality evaluation of education programs (Grudnitski, 1997).

However, everything has its limitations. According to Jae and Cowling (2009), the likelihood of bias in grading will increase if the grader knows the student being assessed. This means using analytic rubric to develop student profile may expose instructors to the so-call "halo effect", where evaluations of a person are impacted by initial concrete impression about the phenomenon that is assessed (Kahneman, 2003; Rosenzweig, 2014). Therefore, criteria and descriptions in the rubric need to be explicitly clarified and carefully designed to reduce the biased and subjective judgments (Andrade & Du, 2005; Murphy, 2004; Neumann & Forsyth, 2008). In addition, clearly identified criteria are also supposed to be useful to ensure high inter-rater consistency and inter-rater agreement, which are the two dimensions of inter-rater reliability (Peeters et al., 2014). The process of converting qualitative terms into a set of scores and overall grade is also realized as a reason for subjectivity and bias in assessment (Cooper & Gargan, 2011). Oermann et al. (2009) warned that without a reliable evaluation tool, issues of inconsistency and subjectivity can undermine the evaluation process leading to grade disputes, course failures, and legal battles. For this reason, the rubric is highly required to pass the test of reliability before being officially applied in practice (Andrade, 2005). Both designers and users may be concerned about the issues of cost and time of developing such kind of rubrics.

CHALLENGES IN DESIGNING, IMPLEMENTING AND APPLYING RUBRICS

BOTH INSTRUCTORS AND LEANERS STILL FACE UNFAVORABLE OUTCOMES AND GRAMMAR CONSISTENCY

Rubrics are educational tools constructed not only to standardize assessment but also stimulate learning and facilitate instruction. However, the world is not always simple. Empirical studies on the impact of using rubrics have revealed the existence of flaws and pitfalls, which may significantly diminish the benefits of using rubrics.

At first glance, scoring with a rubric ensures reliability and validity of judgment. However, Jönsson and Svingby (2007) stated that content validity is the only aspect that might benefit from the use of rubrics while construct validity and criterion validity are almost far from any influence. They argued that "Just by providing a rubric there is no evidence for content representativeness, fidelity of scoring structure to the construct domain or generalizability" (Jönsson & Svingby, 2007). It is even more interesting that reliability of grades is not always improved in all cases. According to G. T. Brown et al.

(2004), consensus estimates, consistency estimates, and measurement estimates, which are three main approaches to determining the accuracy and consistency of scoring, are all required to establish inter-rater accuracy. However, in a pilot test of the New Standard Project on-demand (Table 5), writing tasks scored by 114 teachers, using a seven-point scale holistic rubric results in only 0.54 of consistency coefficient on average, and exact agreement percentages fall in the interval of 40%-49% (Resnick et al., 1993). When comparing analytic and holistic six-point scoring rubrics of Grade 3 writing, Gearhart et al. (1992), found the lowest level of exact agreement percentages was only 15%. To compare six-point narrative-specific and general analytic scoring rubrics, Herman et al. (1993) found the consensus and exact agreement was 39%–46% and 28%–37%, respectively. Meanwhile, average consistency correlations were later estimated from 0.48 to 0.68 (Gearhart et al., 1994). Measurement coefficients when scoring with two different six-point scoring rubrics were reported ranging from 0.47 to 0.68. Those results are also in line with Koretz et al. (1992), Longford (1994), Stuhlmann et al. (1999) and Penny et al.'s (2000) findings. Hunter and Docherty (2011) recognize that even when employing the same rubric, discrepancies and inconsistencies still exist among assessors. In practice, rubrics may cause more problems in writing the assessment than provide solutions (Boulet et al., 2004; Schenck & Daly, 2012). Instructors, thereby, are still discreet with questions related to applying rubrics as assessment and instruction tools especially in writing classes (Broad, 2003). Crusan (2015, p. 2) writes: "I worry about allowing criteria in my head to shape my assessments; I worry about failing to inform students of assignment criteria; I worry about neglecting to provide students with some sort of formalized scale to use as a guide. If rubrics are not used to assess writing, assignment criteria must be made available; some sort of explanation of criteria that allows students to know what will be assessed and how it will be assessed is necessary. Without that, what is the alternative?"

Sometimes the role of rubrics in improving student's performance is questioned. In an attempt to examine the hypothesis that enhanced teacher knowledge of generic assessment rubrics results in improved student achievement (Shafer et al., 2001), the authors concluded that although it seemed there was no detrimental effect on student achievement, there was no evidence found to support instructional value of teacher knowledge of rubrics in neither English nor in government courses. At least 43.3% of students in Laurian and Fitzgerald's (2013) sample functioned as well as or better when using no rubrics. In the same vein, positive effects are identified in only some areas (Green & Bowser, 2006; Sadler & Good, 2006; Schafer et al., 2001; Schamber & Mahoney, 2006; Schirmer et al., 1999), or only when they are combined with other interventions (Toth et al., 2002), or even no differences in the quality of the work are confirmed (Green & Bowser, 2006). For example, Andrade (2001) found that simply providing a rubric to students resulted in better writing in only one out of three essays. In accordance with Andrade (2001), rubric-provided students in the sample of Covill's (2012) study did not write better papers than those who were simply required to identify strengths and weaknesses of their own papers. Howell (2014) urged that caution should be exercised in the use of grading rubrics. Although a grading rubric is supposed to be a useful predictor of assignment outcomes, it seems less powerful than baseline course knowledge and overall academic performance prior to course start (Howell, 2014). In addition, scripts are supposed to be more effective than rubrics at improving self-regulation, learning and self-efficacy (Panadero, 2012). Standardizing the scoring process is not devoid of limitation if it also standardizes prose in the discipline of literature (Kohn, 2006). As insinuated by Balester (2012), this standardization may have a gatekeeping effect, which may promote a focus on errors. Meanwhile, Weigle (2007) claimed that rubrics limit the definition of good writing. Considering all these shreds of evidence of imperfectness, which can also be more extended, the important question that needs to be systematically answered is "What are the reasons for all those unpromising results?"

Table 5. Accuracy statistics of various studies related to the scoring of writing

		Consensus estimates	*Consistency estimates*	*Measurement estimates*
Resnick et al. (1993)	7-point rubric	exact agreement percentages: 40% - 49% (*) adjacent scoring percentages: 86% - 88%	average consistency coefficient: 0.54 (*)	-
Gearhart et al. (1992)	Analytic and holis-tic 6-point scoring rubrics	exact agreement percentages: 15% - 97% (*); adjacent agreement percentages: 80% - 100%	average Pearson consistency coefficient: 0.70	-
Gearhart et al. (1994)	6-point narrative-specific ru-brics	consensus exact agreement: 39%–46% (*); adjacent agreement: 95%–97%	average consistency correlations: 0.48 - 0.68 (*)	measurement coefficient: 0.47 - 0.68 (*)
	6-point narrative general an-alytic scor-ing rubrics	consensus exact agreement: 28%–37% (*); adjacent agreement: 92%–94%	average consistency correlations: 0.48 - 0.68 (*)	measurement coefficient: 0.47 - 0.68 (*)
Novak et al. (1996)	holistic 6-point ru-bric	-	-	dependability coefficient: 0.75 (*)
Koretz et al. (1992)	4-point scoring ru-bric	exact agreement percentages: 44% - 48% (*)	-	-
Supovitz et al. (1997)	8-point ru-bric	exact agreement consensus: 63% - 72% (*)	consistency coeffi-cient: 0.68 - 0.73 (*)	-
	6-point ru-bric	adjacent agreement: 87%–98%	consistency coeffi-cient: 0.75 - 0.87	-
	6-point holistic ru-bric	consensus exact agreement: 59% - 63% (*)	consistency coeffi-cient: 0.74 - 0.75	measurement coefficient: 0.72 - 0.85 (*)
Stuhlmann et al. (1999)	6-category, four-point rubric	-	-	measurement coefficient: 0.61 - 0.66 (*)
Lee (2001)	6-point scale	adjacent agreement consensus: 96%	alpha consistency coefficient: 0.84	-

Source: G. T. Brown et al. (2004)

WHAT KEEPS RUBRICS FAR FROM OPTIMAL?

THE USERS OF RUBRICS

The evidence and findings above are by no means a denial and nor able to implicitly reject the positive contributions of rubrics. They, instead, impart the fact there is a possibility that rubrics may be inefficient and not fulfil their role under some unfavorable circumstances. Thus, identifying factors that determine the probability of that unsatisfactory performance should not be intentionally disregarded.

It is quite redundant to enumerate the benefits of rubrics for learners, instructors and assessors. However, rubrics are not magic wands that can turn dust into gold by themselves. As a supportive instrument, the effectiveness of rubrics undoubtedly depends on the users. In order to take full advantage of rubrics for self-regulated learning, students are assumed to be more likely to use self-tests and self-questioning as sources of feedback to correct misconceptions and to redirect the use of learning strategies, which are not found in novice learners (A.L. Brown et al., 1983; Rafoth et al., 1993; Stein et al., 1982), who "rarely reflect on their own performances and seldom evaluate or adjust their cognitive functioning to meet changing task demands or to correct unsuccessful performances" (Ertmer & Newby, 1996, p. 6). When observing the impact of instructional rubrics on students' writing quality, Andrade (2001) found inconsistent results. He concluded that while instructional rubrics can inform students of the assignment criterion, results may be influenced by several factors including unfamiliarity with using rubrics. Likewise, students in the sample of Covill (2012) did not write better papers with rubrics than those who were simply required to identify strengths and weaknesses of their own papers. Andrade (2001) corroborated that when students are not regularly exposed to the utilization of rubrics, familiarity and effectiveness may be diminished. Orsmond et al. (1996) argued that when students felt unqualified to interpret criteria, they might be reluctant to find the qualities in their work even if they knew what to look for.

Rubrics are even supposed to have little educational value to a number of college-level and university-level teachers (Hafner & Hafner, 2003). The authors accuse academic appointment outside the education department and minimal preparation in teacher education of the unfamiliarity with the popular pedagogical trends including using rubrics as teaching and assessment resources. In the same line, Rezaei and Lovorn (2010) found that in the absence of understanding about effective rubric creation and use, the improvement of reliability or validity seems unlikely. Therefore, training and heightening perceptions of both instructors and learners about rubrics are unavoidable challenges (Martens, 2018). The training sessions, according to Timmerman et al. (2011), significantly increase consistency and reliability of most scores. While instructors need to thoroughly master how to design and employ rubrics effectively, rubrics should not be perceived by the learners, as the real moon but instead they are fingers pointing at the moon. In other words, learners must not prevent themselves from learning further what is mentioned in rubrics. Young (2009) argued that without fully understanding the expectations for each writing assignment, students may be confused and focus only on their final grade. And according to Timmerman et al. (2011, p. 25) "Instructors and program evaluators are cautioned to view the rubric as a tool rather than an answer". After all, a rubric is only a pedagogical tool and it is the users who determine its effectiveness.

CHALLENGES IN DESIGNING RUBRICS

When Tierney and Simon (2004) warned that negative languages and dichotomous tones in some rubrics might imply wrong messages and destroy learners' motivation, educators were once again reminded about an undeniable fact that the design of a rubric strongly relates to how effective and efficient it is. Simply put, rubrics are required to be tailored to specific contexts for specific purposes (Crusan, 2015). Therefore, the content and structure of rubrics should also be seriously put under scrutiny when the learning performance or reliability and validity of assessment fail to meet expectations.

A well-built rubric is one that closely aligns with the objectives of learning, instruction and assessment. Thus, the first step in designing a rubric should be clarifying the goals and targets of activities. When objectives are not clearly defined, consistency and reliability of evaluation may noticeably decrease (Shipman et al., 2012). Meanwhile, teachers hardly find it easy to choose instructional approaches and students are unlikely to be able to achieve a learning expectation if the target is still beyond their understanding (Arter & McTigue, 2001; Stiggins, 2001). "Goals" are broad statements of expected student outcomes and "objectives" divide the goals into observable behaviors (Rogers & Sando, 1996). Ideally, both goals and objectives should convey information and focus that is important for students, instructors as well as assessors (Moskal, 2003). Those objectives, which are also called performance criteria, must indicate all crucial aspects of the goals and describe measurable student outcomes (Moskal, 2003). Interestingly, according to Timmerman et al. (2011, p. 526), "there was a noticeable relationship between the degree to which rubric criteria were included in the course assignment and the reliability of scores generated for that criterion". Simply put, including and omitting criteria in a rubric might put the designers into a dilemma. Lack of relevant criteria prevents students from identifying desired performance skills and consequently, those skills are not improved over time (Fraile et al., 2017). Additionally, the absence of crucial performance criteria is accused of causing low-reliability scores (Timmerman et al., 2011). However, Timmerman et al. (2011) also found an increase in the average student score when criteria were omitted. The suggestion is that criteria must be clearly articulated within a rubric and should reflect the critical underlying characteristics of performance (Tierney and Simon, 2004).

Furthermore, the rating scale of rubric criteria is deemed to be a key factor influencing the rubric's effectiveness. Based on the Rasch vertical rulers and the category response curves, Janssen et al. (2015) provided a visual demonstration that including too many scores for each category scale does not make scores meaningfully different from each other. In an earlier study, it was uncovered that people had difficulty reliably distinguishing between more than about seven levels (Miller, 1956). Nevertheless, a rating scale without enough ranks may also be problematic. Although a smaller number of levels leads to more decision power (Peeters et al., 2014), the fewer scoring bands that appear in the scale, the less powerful the rubric is in terms of distinguishing students' performance because each level of mastery now embraces a wide range of marks. As a rule of thumb, Janssen et al. (2015) recommended a score with 4 to 6 levels to ensure there was distinction without losing meaningful differences. Davidson (2004) offered a similar suggestion that six categories "tend to be workable in most cases". Meanwhile, Andrade and Du (2005) suggested that by answering the question "What do the best and worst performances look like with regard to quality and success?" in advance, the middle levels of the rubric are then more easily filled in. Systematic revision of the scale is deemed to be necessary so that ambiguous or superfluous levels are detected and removed (Crusan, 2015). In order to maintain the category function of rubrics without sacrificing the reliability and simplicity of scoring, rubric designers need to keep up the continuity in the degree of difference between levels and avoid fuzzy descriptions of each scoring band (Dickinson & Adams, 2017; Moskal, 2003). Poor choice of words for criteria and level descriptions may lead to undesirable implications. Using negative language to describe low levels of performance criteria will end up demotivating students (Tierney and Simon, 2004). On the other hand, it is widely supposed that ambiguous language hinders the users from accurately or consistently interpreting the criteria (Beyreli & Ari, 2009; Fang & Wang, 2011; Payne, 2003; Schenck & Daly, 2012). If the descriptions are perceived as abstract and vague by the students, rubrics are not likely to be valuable in promoting learning (Li & Lindsey, 2015) while assessors will assign grades based on their overall impression rather than the criteria described (Knoch, 2009; Weigle, 2002). Therefore, precise and descriptive language; tangible, qualitative terms; indicators and exemplars are highly recommended to generate descriptions of attributes and performance criteria in rubrics (Bargainnier, 2003; Dickinson & Adams, 2017; Payne, 2003; Tierney & Simon, 2004). Ironically, precise language and expanding explanatory details in rubric criteria is supposed to make subjectivity more visible in some other studies (J. Huang, 2012; Turley & Gallagher, 2008).

Consequently, this non-formulary nature of wording makes the process of writing descriptions become a time-consuming task and perhaps the most challenging aspect of rubric design (Moni et al., 2005; Tierney & Simon, 2004). Making decisions on the specificity degree of rubric and scoring strategy is also a troublesome job. It is more than a binary option between generic and task-specific rubrics (Dawson, 2017). Similar to over-fitting prediction models, over-specific rubrics are suspected of narrowing the perception about good and not-good-enough performance, which may diminish students' creativity (Balester, 2012). In regard to the analytic scoring strategies, Tedick, (2002) wrote "Separate scores for different aspects of a student's writing or speaking performance may be considered artificial in that it does not give the teacher (or student) a good assessment of the "whole" of a performance", while a more holistic one is less valuable in giving meaningful feedback.

In practice, there is no chance to design a so-called well-fitted rubric without trustworthy information about the users, who are either instructors or learners. Variables such as personality, age, gender, field of study, types of assignment, education background, educational level and learning orientation of students have been proven to potentially impact learning behaviors and thereby the way rubric is used (Andrade et al., 2009; G.T. Brown et al., 2004; Panadero and Jönsson, 2013). According to Andrade et al. (2009, p. 296), "girls may be affected by rubric-referenced self-assessment in a way that boys are not" due to the fact that girls tend to be more concerned with mastering a writing task than do boys, who, on average, tend to be more concerned with showing someone else that they are capable. In addition, class size and teaching style are important independent variables to be considered to design an effective rubric (Simon & Forgette-Giroux, 2001). From a statistical point of view, valid information must be extracted from a reliable data set. It is noteworthy in a PhD that there are three major factors that cause error and reduce reliability, namely, observer error, environmental changes, and participant changes (*Threats to reliability*, 2012). These are inherent pitfalls in collecting and interpreting data that analysts must encounter and overcome.

In the interest of brevity, when it comes to the commission of designing an efficient rubric, there are two fundamental questions rubric designers need to answer: who the target users are and how to optimize the formation of the rubric. However, things are never as simple as this. Dawson (2017), in an effort to provide a synthesis of the diversity of rubrics, summarized fourteen essential elements (see Table 6) that need to be specified in order to create a complete rubric as well as successfully implement it in classroom. This means designing a so-called well-built rubric is like optimizing a multivariate function. The problem is that the effectiveness of using rubrics non-monotonically depends on its exogenous variables. Besides, ranking the efficacy of potential rubrics of a given course is often an opinion-based decision. The measurements of the goodness of a rubric are indeed recorded after the course ends.

Table 6. Summary of the rubric design elements

Design element	Definition
Specificity	The particular object of assessment
Secrecy	Who the rubric is shared with, and when it is shared
Exemplars	Work samples provided to illustrate quality
Scoring strategy	Procedures used to arrive at marks and grades
Evaluative criteria	Overall attributes required of the student
Quality levels	The number and type of levels of quality
Quality definition	Explanations of attributes of different levels of quality
Judgement complexity	The evaluative expertise required of users of the rubric
Users and uses	Who makes use of the rubric, and to what end
Creators	The designers of the rubric

Design element	Definition
Quality processes	Approaches to ensure the reliability and validity of the rubric
Accompanying feedback information	Comments, annotation, or other notes on student performance
Presentation	How the information in the rubric is displayed
Explanation	Instructions or other additional information provided to users

Source: Dawson (2017)

THE PARADOX OF RUBRIC IMPLEMENTATION

Last but not least, while benefits and challenges of using rubrics have still been openly debated, it is rare to find studies with concrete evidence to reject the fact that students and instructors' perception about rubrics and using rubrics are indeed heterogeneous. As an example, according to the study of Li and Lindsey (2015) on 119 university students and 5 instructors in 2014 in the United States, rubrics were recognized as an instructional tool by most students while the instructors used rubrics primarily for assessment purpose. This is by no means a new finding except the study also clarified how differently rubrics are used by those users. Li and Lindsey (2015) noted that their students tend to focus on keywords and points that are regarded as less important by the instructors. Moreover, instructors show better consistency in interpretation of rubric criteria and descriptions while several students may not completely understand important concepts. Greater variance of students' rating, on the other hand, demonstrated the existence of differences in evaluation between students and instructors. Another example is from Lindsey and Crusan's (2011) study, which also underlined an argument that the rater's assumptions about the writer's ethnicity might cause individual to raters apply rubrics differently.

The existence of discrepancies between students and instructors pose an unwilling paradox: rubrics are supposed to be fully effective only if they are well tailored to the target user with specific context and goal but in practice, both students and instructors commonly share the same rubric for learning, instruction and assessment purposes. This issue should be considered as the knottiest challenges of designing and implementing rubrics in education. Involving students in the process of designing rubrics and providing initial training are promising solutions to homogenize the understanding about rubrics among users (Deci & Ryan, 2000; Vansteenkiste et al., 2006). However, according to Stevens and Levi (2005) educators should not intentionally disregard the time constraints.

RUBRICS AND AUTHENTIC ASSESSMENT

AUTHENTIC ASSESSMENT

Educational assessment functions as the bridge between learning and teaching gap. However, the gap between learning and working is even a thornier issue and receives attention from both educators and employers. Graduate Careers Australia (GCA, 2016) reported that in 2015, only 74% of university graduates had found a full-time job within four months of graduation. The figure evidences a downtrend from 82% to 80% and 76% in 2009, 2010, and 2012 respectively (GCA 2016). Furthermore, in a survey from Harris-Chegg Foundation (2013) on 2001 college students in the United States and 1000 hiring managers, 70% of students seemed to be overconfident when they scored themselves as effective communicators in this area as opposed to 44% of their employers. Authentic learning activities, which emphasize on developing skills, knowledge and attitude required in real-life practice, are key solutions to this learning-working gap (Gulikers et al. 2008; James & Casidy 2018; Tout et al., 2014). According to the principle of learning, instruction and assessment alignment, authentic learning needs to be operated along with authentic assessment.

Authentic assessment is by no means a new concept. Most definitions offered for authentic assessment can be viewed as scoring-centered, context-centered or student-role-centered. According to Savery and Duffy (1995), authentic assessment is based on a criterion situation that reflects or simulates a real-life situation that could confront students in their future professional life. Cumming and Maxwell (1999) considered authentic assessment as an attempt to understand learning complexities by clarifying the relationship between knowledge and interaction in social context. Other arguments emphasized on the realistic value of the tasks in authentic assessment (Herrington & Herrington, 1998; Resnick et al., 1993; Wiggins, 1993). Wiggins (1993, p. 206) noted that "tasks are either replicas of or analogous to the kinds of problems faced by adult citizens and consumers or professionals in the field". Newmann (1996, p. 361) described authenticity as "the extent to which a lesson, assessment task, or sample of student performance represents construction of knowledge through the use of disciplined inquiry that has some value or meaning beyond success in school". It means that student's abilities are judged through real-world-replicated activities (Douchy, 2001; Svinicki, 2004). Authentic assessment is even interchangeably used with performance assessment in some other studies (Hart, 1994; Torrance, 1995). Kirst (1991, p. 21) wrote "use of the term authentic assessment is intended to convey that the assessment tasks themselves are real instances of extended criterion performances, rather than proxies or estimators of actual learning goals. Other synonyms are direct or performance assessments". Others described authentic assessment by what it is not (e.g. not a norm-reference, not employ standardized tests) (Tanner, 2001).

Despite the variety of definitions about authentic assessment, none of them seems to be devoid of drawbacks. Firstly, all scoring-centered, context-centered and student-role-centered definitions fail to fully depict other essential features of authentic assessment. Secondly, although rote learning, for example, supposedly frequently appears in other traditional assessment process, if the real situation demands rote learning, the assessment should also involve learning by rote. Finally, opposing views about the relation between performance and authentic assessment, which argue if authentic assessment is a subset of performance assessment or vice versa (Frey et al., 2012), evidently imply authentic assessment cannot be identified as performance assessment.

In view of these points, the nature of authentic assessment as well as the characteristics of criterion situations seem to be more sufficiently described by Gulikers et al. (2004, p. 69): authentic assessment is "an assessment requiring students to use the same competencies, or combinations of knowledge, skills, and attitudes, that they need to apply in the criterion situation in professional life". Moreover, in accordance with the five-dimension framework for designing authentic assessment, the degree of authenticity is determined by the resemblance between assessment tasks, the physical context, the social context, the assessment result, the assessment criteria, and the real-life criterion situation (Gulikers et al., 2004). This concept and framework allow educators not only better capture the meaning of authentic assessment but also examine and measure the level of authenticity.

THE IMPACT OF AUTHENTIC ASSESSMENT ON LEARNING AND TEACHING

In contrast with the testing culture, which primarily consists of decontextualized designed items in a choice-response format (Birenbaum & Dochy, 2012), problem tasks in authentic assessment more truly reflect knowledge, skills and attitudes, which are required in professional life (Kirschner et al., 2004; Van Merriënboer, 1997). Students' skills and abilities are often assessed under specific physical and social context; tools, resources and time constrains of professional practice (Herrington & Oliver, 2000; Wiggins, 1989). Furthermore, criterion-referenced judgment is one of the most typical features distinguishing authentic assessment with the others. These modifications in assessment approach bring positive stimulations to both students and instructors.

Educators implement authentic assessment with the primary aim of improving students' competencies and employability. Results from empirical studies have not disappointed the advocators. Employability is understood as the "possession of basic "core-skills" or an extended set of generic attributes, or attributes that a type of employer (discipline-linked, sector-related, company-type) specifies"

(Harvey, 2001, p. 96). Oral and written communication skills are important criteria highly sought after by employers across different professions (Chan, 2011; Gray et al., 2005). Moreover, graduate knowledge of the relevant profession, the ability to interact with the profession and seek a career are supposed to influence work-readiness (Jackson, 2016). Authentic assessment exposes students to a wide range of skills such as research, writing, revising, oral skills, debating, and other critical thinking skills (Wiggins, 1990). According to Sotiriadou et al. (2019), interactive oral assessment as authentic assessment design significantly enables the achievement of student engagement, employability, and academic integrity. By observing 120 undergraduate-business students, James and Casidy (2018) realized that authentic assessment improves student satisfaction and promotes positive behavior.

Additionally, positive influences of authentic assessment on instructors' work are also realized. Firstly, authentic assessment is a criteria-based approach that evaluates performance or products against certain criteria. Therefore, the reliability of assessment is considerably increased in comparison with norm-referenced assessment. Secondly, because real-life situations are carefully replicated in authentic tasks, authentic assessment provides high validity, allow both evaluators and students measure the true depth of learning and understanding. Finally, the variety of assessment tools and forms facilitate instructors to implement authentic assessment across many different disciplines and contexts namely law (Martens et al., 2007), nursing (Gulikers et al., 2004), social work (Gulikers et al., 2006), education (Kearney, 2013), and business (James & Casidy 2018; Sotiriadou et al., 2019).

For brevity, when the fact that students do not usually spend time on non-assessed academic work is hardly rejected (Nasab, 2015), authentic assessment tasks, which emphasize knowledge construction, complex thinking, elaborated communication, collaboration and problem solving in authentic contexts, will equip students with vital working skills beyond their classrooms. Meanwhile, a criteria-referenced approach is the key to ensure high reliable and valid of authentic assessment.

THE ROLE OF RUBRICS IN THE IMPLEMENTATION OF AUTHENTIC ASSESSMENT

Although authentic assessment has proven to be effective in improving students' learning and assessment by providing a true depth of student understanding, academic growth, competencies and employability reliability and validity remain important challenges to advocators (Messick, 1996; Rhodes & Finley, 2013). One of the fundamental requirements to implement authentic assessment successfully is providing students scoring criteria with explicit descriptions beforehand, and a rubric seems to be an ideal tool. A rubric as an assessment tool to promote higher-order cognitive skills (Newmann et al., 1996), allows students control over the conditions and contexts they will be assessed, as well as give instructors an anchor to design appropriate instructional strategies (Archbald, 1991; Tellez, 1996). Scoring rubrics, in addition, are widely supposed to be a vital instrument in the implementation of authentic assessments for reliably and validly assessing outcomes that represent workplace tasks.

According to Jönsson and Svingby (2007), consistency of judgement and scoring across students, tasks, and different assessors is widely recorded to be significantly improved when scoring rubrics are employed in authentic and performance-based assessments. The more consistent the scores, the more reliable the assessment is supposed to be. Moskal and Leydens (2000) claimed that consistency is obtained if and only if the whole process of assessment is independent of scorers, time period between the assessments, and the contexts of assessment. Vendlinski et al. (2002) achieved valid inferences of a student's content understanding when authentically assessed 134 first-year high school chemistry students without gender, ethnic, or socioeconomic bias thanks to scoring with rubrics. In addition, Diller and Phelps (2008) advocated the use of rubric as an assessment instrument to enhance the precision of interpreting students' results. Different roles of rubrics in the implementation of authentic assessment are briefly summarized in Table 7.

As a trade-off for representing real-world tasks in educational settings, authentic assessment is exposed to the threats of subjective scoring and judgments (Messick, 1996). Ideally, rubrics with specific evaluating guidelines will probably improve the objectivity and consistency of scoring; ensure higher level of reliability and validity of assessment. However, many extraneous factors that determine the reliability and validity have not been fully addressed (Ghosh et al., 2016) and therefore, the question of increasing the validity and reliability of rubrics as an authentic assessment instrument is left undiscovered.

Table 7. The roles of rubric when used as an authentic assessment instrument

Author/year	Context of study	Type of validity demonstrated	Type of reliability demonstrated	Techniques/ coefficients for validity/reliability of rubrics	Reason for using scoring rubrics
Emery (2001)	School students	-	-	-	Improving student performance through scoring rubrics
Wellington et al. (2002)	School students	-	-	-	To provide a correlation between different measures of student understanding
Moon et al. (2005)	School students	Content validity	Inter-rater reliability	Kappa formula	To provide quantifiable information about student learning and instruction process
Y. L. Johnson (2007)	School students	(1) Face validity (2) Content validity (3) Content relevance	Internal consistency reliability	**Validity:** field experts **reliability:** KR20	To compare student achievement scores on authentic assessment with that on traditional assessments
Olfos and Zulantay (2007)	School students	Concurrent validity	Internal consistency reliability	**Validity:** criteria of judges, parallel instruments, non-obstructive data **reliability:** Rho of spearman, index r of Pearson, Cronbach's alpha	To improve the validity and reliability of the web-based authentic assessment system
Jönsson (2008)	University students	(1) Face validity (2) Construct validity	(1) Internal consistency reliability	**Face validity:** student interviews	To assess student performance and self-

Author/year	Context of study	Type of validity demonstrated	Type of reliability demonstrated	Techniques/ coefficients for validity/reliability of rubrics	Reason for using scoring rubrics
	(teacher education)		(2) Inter-rater reliability (3) Rank Correlation	**Content validity**: experts' validation Internal **consistency reliability**: Cronbach's Spearman's rho Rank Correlation: Pearson's r	assessment skills of students in authentic assessment
Diller and Phelps (2008)	University program	Validity demonstrated through reliability tests	Internal consistency reliability	Multivariate; item correlation; factor analysis; Cronbach's alpha	To assess the effectiveness of the course program through authentic assessment
Cassidy (2009)	Elementary school teachers	Convergent validity	Internal consistency reliability	Validity established based on previous use; reliability provided through multiple assessment tasks	To measure relationship between teacher effectiveness (in terms of level of instructional quality) and student achievement through authentic assessment scores
Taylor (2011)	School students	Internal validity	Inter-rater reliability	Threats to internal validity minimized through teacher development workshops, feedback from parents and students; reliability obtained through multiple raters	To measure achievement of learning objectives through interdisciplinary authentic assessment
Azim and Khan (2012)	School students	-	-	-	To assess students' knowledge, higher-order skills, and performance through authentic assessment

Author/year	Context of study	Type of validity demonstrated	Type of reliability demonstrated	Techniques/ coefficients for validity/reliability of rubrics	Reason for using scoring rubrics
Mccarthy (2013)	University students (business graduates)	-	-	-	To use as a self-assessment and feedback tool by articulating lecturer's expectations from students
Blackburn and Kelsey (2013)	School students	-	-	-	To assess student performance in authentic assessment
Fatonah et al. (2013)	School students	Content validity	(1) inter-rater reliability (2) instrument reliability	**Validity**: field experts using Aikends validity **Inter-rater reliability**: Kappa formula **Instrument reliability**: Alpha formula, factor analysis using SPSS and Lisrel	To access student performance in a proposed authentic assessment model
Hensel and Stanley (2014)	University students (nursing education)	-	Inter-rater reliability	Achievement of reliability implied text; empirical measures and data not available	To score student performance in a stimulated authentic assessment task

Source: Ghosh et al. (2016)

ANALYSIS OF ARTICLES REVIEWED

Table 8 provides a well-defined picture of the lowest and highest numbers of publications from all 105 journals used in this review. At least 77% of the journals are reported and ranked by ScimagoJR, 61% of the journals stand in the top 25% of journals for at least one of its classified sub-disciplines (see Figure 1). This highlights the involvement of quality journals in education and related disciplines. Table 8 also indicates that scientific journals specializing in educational assessment hold a major proportion. *Assessing Writing* and *Assessment & Evaluation in Higher Education* contribute the highest number of papers in this review with 9 and 8 articles respectively, making up more than 11% totally. They are followed by *Practical Assessment, Research & Evaluation* (5), *Educational Assessment* (3), *Educational Researcher* (3), and *Educational Technology Research and Development* (3).

Table 8. Journal publications

Name of Journals	Number of Publications	%
Assessing Writing	9	5,96%
Assessment & Evaluation in Higher Education	8	5,30%
Practical Assessment, Research & Evaluation	5	3,31%
Educational Assessment	3	1,99%
Educational Researcher	3	1,99%
Educational Technology Research and Development	3	1,99%
English Journal	3	1,99%
Advances in Health Sciences Education	2	1,32%
Assessment in Education: Principles, Policy & Practice	2	1,32%
Educational Leadership	2	1,32%
Evaluation and Program Planning	2	1,32%
Higher Education Research & Development	2	1,32%
Instructional Science	2	1,32%
Journal of Experimental Psychology: General	2	1,32%
Journal of Nursing Education	2	1,32%
Language Testing	2	1,32%
Nurse Education Today	2	1,32%
Nursing Education Perspectives	2	1,32%
Procedia-Social and Behavioral Sciences	2	1,32%
Review of Educational Research	2	1,32%
Studies in Educational Evaluation	2	1,32%
Studies in Higher Education	2	1,32%
Academic Research International	1	0,66%
Across the Disciplines: A Journal of Language, Learning, and Academic Writing	1	0,66%
Advances in Physiology Education	1	0,66%
American Educational Research Journal	1	0,66%
American Journal of Education	1	0,66%
American Journal of Pharmaceutical Education	1	0,66%
American Psychologist	1	0,66%
Applied Measurement in Education	1	0,66%
Assessment Update	1	0,66%
Astronomy Education Review	1	0,66%
Australian Journal of Language and Literacy	1	0,66%

Name of Journals	Number of Publications	%
British Journal of Educational Technology	1	0,66%
CBE - Life Sciences Education	1	0,66%
College Teaching	1	0,66%
Computers & Education	1	0,66%
Computers in Human Behavior	1	0,66%
Creative Education	1	0,66%
Current Issues in Education	1	0,66%
Currents in Pharmacy Teaching and Learning	1	0,66%
Diagnostique	1	0,66%
Distance Education	1	0,66%
Education and Training in Developmental Disabilities	1	0,66%
Education Sciences	1	0,66%
Educational Horizons	1	0,66%
Educational Psychologist	1	0,66%
Educational Psychology Review	1	0,66%
Educational Research Review	1	0,66%
Educational Sciences: Theory and Practice	1	0,66%
Educational Technology & Society	1	0,66%
Essays on Teaching Excellence	1	0,66%
ETS Research Report Series	1	0,66%
Exceptional Children	1	0,66%
Holist Nursing Practice	1	0,66%
Innovations in Education and Teaching International	1	0,66%
International Journal of Engineering Education	1	0,66%
International Journal of Science Education	1	0,66%
International Journal of Teaching and Learning in Higher Education	1	0,66%
Issues in Informing Science & Information Technology	1	0,66%
Journal for the Study of Education and Development	1	0,66%
Journal of Accounting Education	1	0,66%
Journal of Applied Linguistics and Language Research	1	0,66%
Journal of Chemical Education	1	0,66%
Journal of Education and Practice	1	0,66%
Journal of Educational Technology & Society	1	0,66%
Journal of Effective Teaching	1	0,66%

Name of Journals	Number of Publications	%
Journal of Human Sciences and Extension	1	0,66%
Journal of Information Science	1	0,66%
Journal of Learning Design	1	0,66%
Journal of Library Administration	1	0,66%
Journal of Medical and Dental Sciences	1	0,66%
Journal of Research on Technology in Education	1	0,66%
Journal of Science Education and Technology	1	0,66%
Journal of Second Language Writing	1	0,66%
Journal of Secondary Gifted Education	1	0,66%
Journal of Statistics Education	1	0,66%
Journal of the Scholarship of Teaching and Learning	1	0,66%
Journal of Vocational Education and Training	1	0,66%
Journal of Writing Assessment	1	0,66%
Journal of Excellence in College Teaching	1	0,66%
Learning and Individual Differences	1	0,66%
Learning and Instruction	1	0,66%
Medical Education	1	0,66%
Medical Teacher	1	0,66%
Medicine & Science in Sports & Exercise	1	0,66%
Metacognition and Learning	1	0,66%
Nature Reviews Neuroscience	1	0,66%
Neuroscience	1	0,66%
New Directions for Teaching and Learning	1	0,66%
Portal: Libraries and The Academy	1	0,66%
Practical Assessment, Research and Evaluation	1	0,66%
Psychological Inquiry	1	0,66%
Psychological Review	1	0,66%
Quality in Higher Education	1	0,66%
Reading & Writing Quarterly	1	0,66%
Reading Psychology	1	0,66%
School Psychology Quarterly	1	0,66%
Science	1	0,66%
Science Education	1	0,66%
Teacher Education and Special Education	1	0,66%

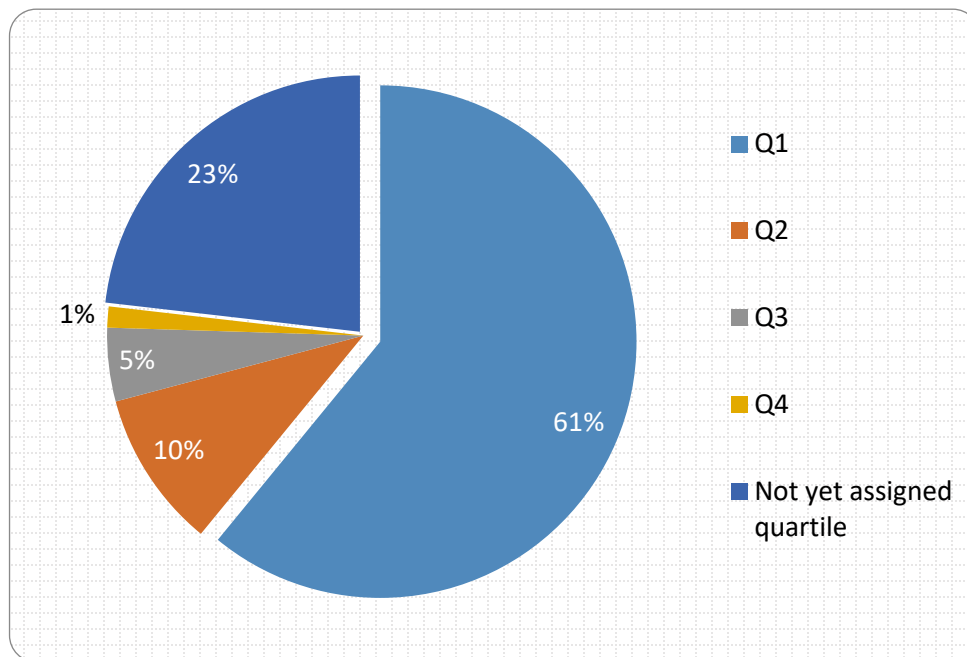| Name of Journals	Number of Publications	%
Teaching and Teacher Education	1	0,66%
Teaching in Higher Education	1	0,66%
Teaching of Psychology	1	0,66%
The High School Journal	1	0,66%
The Journal of Continuing Education in Nursing	1	0,66%
The Journal of Educational Research	1	0,66%
The Journal of General Education	1	0,66%
WMU Journal of Maritime Affairs	1	0,66%
	151	**100%**

Figure 1. Journal ranking

Figure 2 illustrates various formats of publications that are cited in this research. Overall, to ensure high reliability, scientific articles and books are the main sources of reference. They both together accounted for 95% of all formats. A small number of other sources of reference such as doctoral dissertations approved by reputable universities, interim reports from specialized organizations and high-rated online documents are also exploited as supplemental sources of practical aspects.
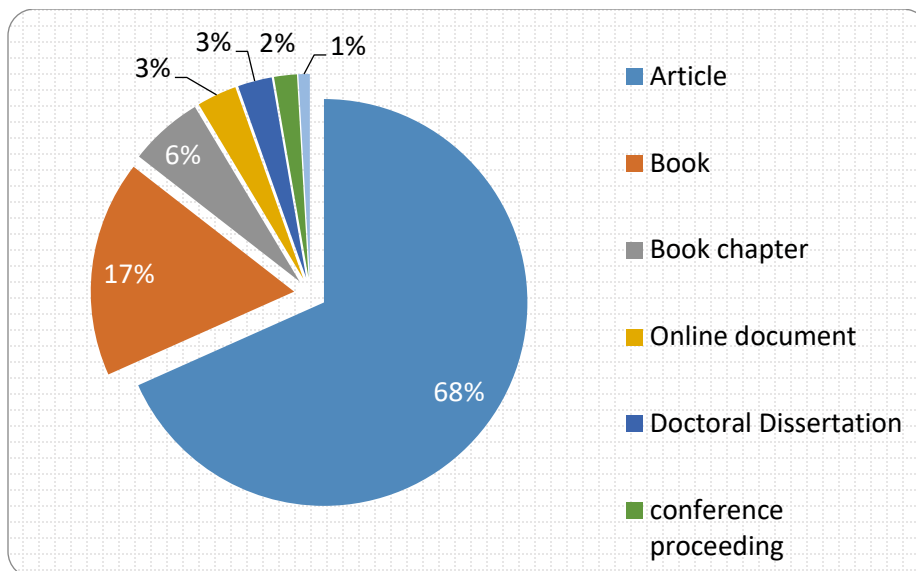
Figure 2. Source of references

CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

There are many studies that have been done on rubric design, benefits and disadvantages associated with rubric usage. This study provides an overall understanding of several types of rubric designs for authentic assessment as well as the benefits and challenges.

A rubric is not only an assessment tool useful for students in high-stakes exam. It is indeed an educational instrument supporting learners to select appropriate learning approaches, assisting teachers to design effective instruction strategies, and improve reliability and validity of assessment.

The variety of rubric formats brings the users more options for different contexts. Novice learners should begin with the generic rubric due to its simplicity. Meanwhile, a task-specific rubric is more useful to improve reliability and validity of large-scale assessment. A holistic rubric is appropriate for assessment of learning and an analytic rubric is almost indispensable in student-centered classroom and assessment for learning.

Rubric users, however, should not forget inherent challenges of each type of rubric. There are three main factors that determine how effective a rubric is in improving students' performance namely the perception of users about rubrics, the design of the rubric and differences in the purpose of using rubrics. Therefore, both students and teachers should be given training sessions to effectively employ rubrics for learning, teaching and assessment. Rubric designing variables should also be optimized based on reliable data and information about the target educational context.

Modern education philosophy places more emphasis on competencies and employability of students. Consequently, students are increasingly encouraged to participate in authentic learning activities. Assessment methods are also modified to ensure the alignment between learning, instruction, and assessment. However, authentic assessment can also be subjective. Using rubric as an authentic assessment instrument has been widely demonstrated to be useful in enhancing the reliability of authentic assessment. Moreover, other empirical results indicate that rubrics play an important role in authentic assessment regardless of levels or disciplines.

This study is rather a general narrative review from literature over the past 15 years of the most important and critical aspects of the current knowledge about rubrics and authentic assessment. Although this review article aimed to provide readers with up-to-date knowledge, there is neither a methodological approach that interprets data nor answers to specific quantitative research questions.

The overall trend, the current focal point as well as potential related fields of study are not clearly illustrated with statistical data. For future research in rubrics and the relevance in authentic assessment, a systemic literature review which collects secondary data, critically appraise research studies, and synthesize findings qualitatively or quantitatively will be a useful supplement.

REFERENCES

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education*, *5(*3), 197-203. https://doi.org/10.1187/cbe.06-06-0168

Alonso-Tapia, J., & Panadero, E. (2010). Effects of self-assessment scripts on self-regulation and learning. *Journal for the Study of Education and Development, 33*(3), 385-397. https://doi.org/10.1174/021037010792215145

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, *57(*5), 13–18. https://doi.org/10.1007/BF01630914

Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, *4*(4). https://cie.asu.edu/ojs/index.php/cieatasu/article/view/1630

Andrade, H. L., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Educational & Counseling Psychology*, *2*. https://scholarsarchive.library.albany.edu/edpsych_fac_scholar/2

Andrade, H. L., Wang, X., Du Y., & Akawi, R. A. (2009). Rubric referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, *102*(4), 287-302, https://doi.org/10.3200/JOER.102.4.287-302

Archbald, D. A. (1991). Authentic assessment: Principles, practices, and issues. *School Psychology Quarterly*, *6*(4), 279. https://doi.org/10.1037/h0088821

Arter, J., & Chappuis, J. (2006). *Creating and recognising quality rubrics*. Pearson Education.

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin Press.

Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education*, *110*, 154-169. https://doi.org/10.1016/j.compedu.2017.03.012

Azim, S., & Khan, M. (2012). Authentic assessment: An instructional tool to enhance students learning. *Academic Research International*, *2*(3), 314.

Balan, A. (2012). *Assessment for learning: A case study in mathematics education*. Malmö University Press.

Balester, V. (2012). How writing rubrics fail: Toward a multicultural model. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 63–77). Peter Lang Publishing.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61(*2), 213-238. https://doi.org/10.3102/00346543061002213

Bargainnier, S. (2003). Fundamentals of rubrics. *Pacific Crest*. http://www.webpages.uidaho.edu/ele/Scholars/Practices/Evaluating_Projects/Resources/Using_Rubrics.pdf

Barkley, E. F., Cross, K. P., & Major, C. H. (2014). *Collaborative learning techniques: A handbook for college faculty*. John Wiley & Sons.

Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, *35*(2), 180-191. https://doi.org/10.1177/0165551508095781

Beyreli, L., & Ari, G. (2009). The use of analytic rubric in the assessment of writing performance – Inter-rater concordance study. *Educational Sciences: Theory and Practice*, *9*(1), 105-125.

Birenbaum, M., & Dochy, F. (Eds.) (2012). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Springer Science & Business Media.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Blackburn, J. J., & Kelsey, K. D. (2013). Understanding authentic assessment in a secondary agricultural mechanics laboratory: An instrumental case study. *Journal of Human Sciences and Extension*, *1*(1), 1-16. http://media.wix.com/ugd//c8fe6e_0d819a82a447c3dd617f45ad348d4358.pdf

Bolton, A. (2006). Immediate versus delayed feedback in simulation-based training: Matching feedback delivery timing to the cognitive demands of the training exercise. *Electronic Theses and Dissertations 2004-2019*. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1925&context=etd

Boulet, J. R., Rebbecchi, T. A., Denton, E. C., McKinley, D. W., & Whelan, G. P. (2004). Assessing the written communication skills of medical school graduates. *Advances in Health Sciences Education*, *9*(1), 47-60. https://doi.org/10.1023/B:AHSE.0000012216.39378.15

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470-497). Cambridge University Press https://doi.org/10.1017/CBO9780511529863.022

Bransford, J. D., Stein, B. S., Vye, N. J., Franks, J. J., Auble, P. M., Mezynski, K. J., & Perfetto, G. A. (1982). Differences in approaches to learning: An overview. *Journal of Experimental Psychology: General*, *111(*4), 390.

Broad, B. (2003). *What we really value: Beyond rubrics in teaching and assessing writing.* University Press of Colorado. https://doi.org/10.2307/j.ctt46nxvm

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading.* ASCD.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. (1983). Learning, understanding, and remembering. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology*, *Volume 3* (pp. 77-167). Wiley.

Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 3-35). Lawrence Erlbaum Associates.

Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education.* Psychology Press.

Brown, G. T., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, *9*(2), 105-121. https://doi.org/10.1016/j.asw.2004.07.001

Brush, T., & Saye, J. (2000). Implementation and evaluation of a student-centered learning unit: A case study. *Educational Technology Research and Development*, *48*(3), 79-100. https://doi.org/10.1007/BF02319859

Busching, B. (1998). Grading inquiry projects. *New Directions for Teaching and Learning*, *74*, 89-96. https://doi.org/10.1002/tl.7409

Cassidy, K. E. (2009). *Using authentic intellectual assessment to determine level of instructional quality of teacher practice of new elementary school teachers based on teacher preparation route* [Doctoral dissertation]. The George Washington University.

Center for Advanced Research on Language Acquisition (CARLA). 2018. *Types of rubrics.* http://carla.umn.edu/assessment/vac/improvement/p_5.html

Chan, V. (2011). Teaching oral communication in undergraduate science: Are we doing enough and doing it right? *Journal of Learning Design*, *4*(3), 71-79. https://doi.org/10.5204/jld.v4i3.82

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise.* Lawrence Erlbaum Associates.

Coe, D. P., Pivarnik, J. M., Womack, C. J., Reeves, M. J., & Malina, R. M. (2006). Effect of physical education and activity levels on academic achievement in children. *Medicine & Science in Sports & Exercise*, *38*(8), 1515-1519. https://doi.org/10.1249/01.mss.0000227537.13175.1b

Cooper, B. B. S., & Gargan, A. (2011). Rubrics in education - Old term, new meanings. *Educational Horizons*, *89*(4), 6-8. https://doi.org/10.1177/0013175X1108900402

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 245-252. https://doi.org/10.1145/365024.365111

Corley, M. (2008). *Student-centered learning.* Adapted from *CALPRO Professional Development Module.* https://lincs.ed.gov/sites/default/files/6%20_TEAL_Student-Centered.pdf

Covill, A. (2012). College students' use of a writing rubric: Effect on quality of writing, self-efficacy, and writing practices. *Journal of Writing Assessment*, *5*(1), 1-19. http://www.journalofwritingassessment.org

Crumly, C., & Dietz, P. (2014). *Pedagogies for student-centered learning: Online and on-ground.* Augsburg Fortress Publishers. https://doi.org/10.2307/j.ctt9m0skc

Crusan, D. J. (2015). Dance, ten; looks, three: Why rubrics matter. *Assessing Writing*, *26*, 1-4. https://doi.org/10.1016/j.asw.2015.08.002

Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policy & Practice*, *6*(2), 177-194. https://doi.org/10.1080/09695949992865

Davidson, E. J. (2004). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* SAGE Publications. https://doi.org/10.4135/9781452230115

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment and Evaluation in Higher Education*, *42*(3), 347-360. https://doi.org/10.1080/02602938.2015.1111294

Deakin University. (2017). *Live the future: Agenda 2020.* https://www.deakin.edu.au/__data/assets/pdf_file/0003/623145/LIVE-2017-2020_Sept.pdf

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry*, *11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Dickinson, P., & Adams, J. (2017). Values in evaluation: The use of rubrics. *Evaluation and Program Planning*, *65*, 113-116. https://doi.org/10.1016/j.evalprogplan.2017.07.005

Diller, K. R., & Phelps, S. F. (2008). Learning outcomes, portfolios, and rubrics, Oh My! Authentic assessment of an information literacy program. *Portal: Libraries and the Academy*, *8*(1), 75-89. https://doi.org/10.1353/pla.2008.0000

Dochy, F. J., & McDowell, L. (1997). Introduction: Assessment as a tool for learning. *Studies in Educational Evaluation*, *23*(4), 279-298. https://doi.org/10.1016/S0191-491X(97)86211-6

El Saadawi, G. M., Tseytlin, E., Legowski, E., Jukic, D., Castine, M., Fine, J., Gormley, R., & Crowley, R. S. (2008). A natural language intelligent tutoring system for training pathologists: Implementation and evaluation. *Advances in Health Sciences Education*, *13*(5), 709-722. https://doi.org/10.1007/s10459-007-9081-3

Emery, D. E. (2001). Authentic assessment in high school science: A classroom perspective. In D. P. Shepardson (Ed.), *Assessment in science* (pp. 227-247). Springer. https://doi.org/10.1007/978-94-010-0802-0_14

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1-24. https://doi.org/10.1007/BF00156001

Ewing, S. C. (1998). Alternative assessment: Popularity, pitfalls, and potential. *Assessment Update*, *10*(1), 1-2. https://doi.org/10.1002/au.101

Fang, Z., & Wang, Z. (2011). Beyond rubrics: Using functional language analysis to evaluate student writing. *Australian Journal of Language and Literacy*, *34*(2), 147.

Fatonah, S., Suyata, P., & Prasetyo, Z. K. (2013). Developing an authentic assessment model in elementary school science teaching. *Journal of Education and Practice*, *4*(13), 50-61.

Feingold, C. E., Cobb, M. D., Givens, R. H., Arnold, J., Keller, J. L., & Joslin, S. (2008). Student perceptions of team learning in nursing education. *Journal of Nursing Education*, *47*(5), 214-222. https://doi.org/10.3928/01484834-20080501-03

Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, *53*, 69-76. https://doi.org/10.1016/j.stueduc.2017.03.003

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109. https://doi.org/10.3102/00346543074001059

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research & Evaluation*, *17*(2).

Gearhart, M., Herman, J. L., Baker, E. L., & Whitaker, A. K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment*. https://files.eric.ed.gov/fulltext/ED344900.pdf

Gearhart, M., Herman, J. L., Novak, J. R., & Wolf, S. A. (1994). Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric. *Assessing Writing*, *2*(2), 207-242. https://doi.org/10.1016/1075-2935(95)90013-6

Ghosh, S., Bowles, M., Ranmuthugala, D., & Brooks, B. (2016). Authentic assessment in seafarer education: Using literature review to investigate its validity and reliability through rubrics. *WMU Journal of Maritime Affairs*, *15*(2), 317-336. https://doi.org/10.1007/s13437-015-0094-0

Gibbs, G. (1992). *Improving the quality of student learning: Based on the Improving Student Learning Project funded by the Council for National Academic Awards*. Technical and Education Services.

Graduate Careers Australia. (2016). Graduate Destinations 2015 – A report on the work and study outcomes of recent higher education graduates. https://www.graduatecareers.com.au/files/wp-content/uploads/2016/07/graduate-destinations-report-2015-final1.pdf

Gray, F. E., Emerson, L., & MacKay, B. (2005). Meeting the demands of the workplace: Science students and written skills. *Journal of Science Education and Technology*, *14*(4), 425-435. https://doi.org/10.1007/s10956-005-8087-y

Green, R., & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, *45*(1-2), 185-202. https://doi.org/10.1300/J111v45n01_10

Grudnitski, G. (1997). A forecast of achievement from student profile data. *Journal of Accounting Education*, *15*(4), 549-558. https://doi.org/10.1016/S0748-5751(97)00024-9

Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67. https://doi.org/10.1007/BF02504676

Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. (2006). Authentic assessment, student and teacher perceptions: The practical value of the five-dimensional framework. *Journal of Vocational Education and Training*, *58*(3), 337-357. https://doi.org/10.1080/13636820600955443

Gulikers, J. T., Bastiaens, T. J., Kirschner P. A. & Kester, L. (2008). Authenticity is in the eye of the beholder: student and teacher perceptions of assessment authenticity. *Journal of Vocational Education & Training*, *60*(4), 401-412, https://doi.org/10.1080/13636820802591830

Gulikers, J. T., Bastiaens, T. J., & Martens, R. L. (2005). The surplus value of an authentic learning environment. *Computers in Human Behavior*, *21*(3), 509-521. https://doi.org/10.1016/j.chb.2004.10.028

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, *25*(12), 1509-1528. https://doi.org/10.1080/0950069022000038268

Halonen, J. S., Bosack, T., Clay, S., McCarthy, M., Dunn, D. S., Hill IV, G. W., McEntarffer, R., Mehrotra, C. Nesmith, R., Weaver, K. A., & Whitlock, K. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, *30*(3), 196-208. https://doi.org/10.1207/S15328023TOP3003_01

Hannafin, M. J., & Hannafin, K. M. (2010). Cognition and student-centered, web-based learning: Issues and implications for research and theory. In J. Spector, D. Ifenthaler, P. Isaias, & S. D. Kinshuk (Eds.) *Learning and instruction in the digital age* (pp. 11-23). https://doi.org/10.1007/978-1-4419-1551-1_2

Harris-Chegg Foundation. (2013). *Bridge that gap: Analyzing the Student Skill Index*. http://www.insidehigh-ered.com/sites/default/server_files/files/Bridge%20That%20Gap-v8.pdf

Harvey, L. (2001). Defining and measuring employability. *Quality in Higher Education*, *7*(2), 97-109. https://doi.org/10.1080/13538320120059990

Hensel, D., & Stanley, L. (2014). Group simulation for "authentic" assessment in a maternal-child lecture course. *Journal of the Scholarship of Teaching and Learning*, *14*(2), 61-70. https://doi.org/10.14434/jo-sotl.v14i2.4081

Herman, J. L. (1992). *A practical guide to alternative assessment*. Association for Supervision and Curriculum Development.

Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, *1*(3), 201-224. https://doi.org/10.1207/s15326977ea0103_2

Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. *Higher Education Research & Development*, *17*(3), 305-322. https://doi.org/10.1080/0729436980170304

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, *48*(3), 23-48. https://doi.org/10.1007/BF02319856

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nature Reviews Neuroscience*, *9*(1), 58. https://doi.org/10.1038/nrn2298

Hillman, C. H., Pontifex, M. B., Raine, L. B., Castelli, D. M., Hall, E. E., & Kramer, A. F. (2009). The effect of acute treadmill walking on cognitive control and academic achievement in preadolescent children. *Neuroscience*, *159*(3), 1044-1054. https://doi.org/10.1016/j.neuroscience.2009.01.057

Honebein, P. C., Duffy, T. M., & Fishman, B. J. (1993). Constructivism and the design of learning environments: Context and authentic activities for learning. In T. M. Duffy, J. Lowyck, D. H. Jonassen, & T. M. Welsh (Eds.) *Designing environments for constructive learning* (pp. 87-108). Springer. https://doi.org/10.1007/978-3-642-78069-1_5

Howell, R. J. (2011). Exploring the impact of grading rubrics on academic performance: Findings from a quasi-experimental, pre-post evaluation. *Journal on Excellence in College Teaching*, *22*(2), 31-49.

Howell, R. J. (2014). Grading rubrics: Hoopla or help? *Innovations in Education and Teaching International*, *51*(4), 400-410. https://doi.org/10.1080/14703297.2013.785252

Huang, H. M. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, *33*(1), 27-37. https://doi.org/10.1111/1467-8535.00236

Huang, J. (2012). Using generalizability theory to examine the accuracy and validity of large-scale ESL writing assessment. *Assessing Writing*, *17*(3), 123-139. https://doi.org/10.1016/j.asw.2011.12.003

Hunter, K., & Docherty, P. (2011). Reducing variation in the assessment of student writing. *Assessment & Evaluation in Higher Education*, *36*(1), 109-124. https://doi.org/10.1080/02602930903215842

Jackson, D. (2016). Re-conceptualizing graduate employability: The importance of pre-professional identity. *Higher Education Research & Development, 35*(5), 925-939. https://doi.org/10.1080/07294360.2016.1139551

Jae, H., & Cowling, J. (2009). Objectivity in grading: The promise of bar codes. *College Teaching*, *57*(1), 51-55. https://doi.org/10.3200/CTCH.57.1.51-55

James, L. T., & Casidy, R. (2018). Authentic assessment in business education: Its effects on student satisfaction and promoting behaviour. *Studies in Higher Education*, *43*(3), 401-415. https://doi.org/10.1080/03075079.2016.1165659

Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, *26*, 51-66. https://doi.org/10.1016/j.asw.2015.07.002

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365-379. https://doi.org/10.3102/0013189X09339057

Johnson, Y. L. (2007). *The efficacy of authentic assessment versus pencil and paper testing in evaluating student achievement in a basic technology course* [Unpublished doctoral dissertation]. Walden University, MN.

Jones, L. (2007). *The student-centered classroom.* Cambridge University Press. https://mail.brettwilkin.com/phoca-download/StudentCentredClassroom/jones-student-centered.pdf

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

Jönsson, A. (2008). *Educative assessment for/ of teacher competency. A study of assessment and learning in the "Interactive examination" for student teachers* [Unpublished doctoral dissertation]. Malmö University, Sweden.

Jönsson, A., & Panadero, E. (2017). The use and design of rubrics to support assessment for learning. In D. Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education. The enabling power of assessment* (pp. 99–111). Springer.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), 697. https://doi.org/10.1037/0003-066X.58.9.697

Karkehabadi, S. (2013). *Using rubrics to measure and enhance student performance.* Northern Virginia Community College. https://www.nvcc.edu/assessment/docs/FTW5.usingrubricsmeasurestuperf-spr13.pdf

Kearney, S. (2013). Improving engagement: the use of 'Authentic self-and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education*, *38*(7), 875-891. https://doi.org/10.1080/02602938.2012.751963

Kirschner, P. A., Martens, R. L., & Strijbos, J. W. (2004). CSCL in higher education? In J.-W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), *What we know about CSCL and implementing it in higher education* (pp. 3-30). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-7921-4_1

Kirst, M. W. (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher*, *20*(2), 21. https://doi.org/10.2307/1176830

Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, *26*(2), 275-304. https://doi.org/10.1177/0265532208101008

Kohn, A. (2006). The trouble with rubrics. *English Journal*, *95*(4), 12-15. https://doi.org/10.2307/30047080

Koretz, D., McCaffrey, D. F., Klein, S. P., Bell, R. M., & Stecher, B. M. (1992). *The reliability of scores from the 1992 Vermont portfolio assessment program: Interim report.* RAND Corporation. https://www.rand.org/pubs/drafts/DRU159.html

Krause, K. L., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, *33*(5), 493-505. https://doi.org/10.1080/02602930701698892

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia – Social and Behavioral Sciences*, *31*, 486-490. https://doi.org/10.1016/j.sbspro.2011.12.091

Laurian, S., & Fitzgerald, C. J. (2013). Effects of using rubrics in a university academic level Romanian literature class. *Procedia – Social and Behavioral Sciences*, *76*, 431-440. https://doi.org/10.1016/j.sbspro.2013.04.141

Learnalberta.ca. (2018). *Generic rubrics.* http://www.learnalberta.ca/content/mewa/html/assessment/genericrubrics.html

Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing*, *26*, 67-79. https://doi.org/10.1016/j.asw.2015.07.003

Lindsey, P., & Crusan, D. J. (2011). How faculty attitudes and expectations toward student nationality affect writing assessment. *Across the Disciplines: A Journal of Language, Learning, and Academic Writing*, 8.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, *20*(8), 15-21. https://doi.org/10.3102/0013189X020008015

Longford, N. T. (1994). A case for adjusting subjectively rated scores in the advanced placement tests. *ETS Research Report Series*, *1994 No.* 2, pp. i-21. https://doi.org/10.1002/j.2333-8504.1994.tb01632.x

Martens, K. S. (2018). How program evaluators use and learn to use rubrics to make evaluative reasoning explicit. *Evaluation and Program Planning*, 69, 25-32. https://doi.org/10.1016/j.evalprogplan.2018.03.006

Martens, R., Bastiaens, T., & Kirschner, P. A. (2007). New learning design in distance education: The impact on student perception and motivation. *Distance Education*, *28*(1), 81-93. https://doi.org/10.1080/01587910701305327

McAlpine, M. (2002). *Principles of assessment*. CAA Centre, University of Luton.

McCarthy, G. (2013). Authentic assessment: key to learning. In E. Doyle, P. Buckley & C. Carroll (Eds.), *Innovative business school teaching: Engaging the millennial generation* (pp. 81-92). Routledge.

McCombs, B. L., & Whisler, J. S. (1997). *The learner-centered classroom and school: Strategies for increasing student motivation and achievement*. Jossey-Bass.

McNamara, T. F. (1996). *Measuring second language performance*. Addison Wesley Longman.

Menéndez-Varela, J. L., & Gregori-Giralt, E. (2016). The contribution of rubrics to the validity of performance assessment: A study of the conservation–restoration and design undergraduate degrees. *Assessment & Evaluation in Higher Education*, *41*(2), 228-244. https://doi.org/10.1080/02602938.2014.998169

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, *7*(25), 1-10.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241-256. https://doi.org/10.1177/026553229601300302

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81. https://doi.org/10.1037/h0043158

Moni, R. W., Beswick, E., & Moni, K. B. (2005). Using student feedback to construct an assessment rubric for a concept map in physiology. *Advances in Physiology Education*, *29*(4), 197-203. https://doi.org/10.1152/advan.00066.2004

Moon, T. R., Brighton, C. M., Callahan, C. M., & Robinson, A. (2005). Development of authentic assessments for the middle school classroom. *Journal of Secondary Gifted Education*, *16*(2-3), 119-133. https://doi.org/10.4219/jsge-2005-477

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*(1-2), 99-113. https://doi.org/10.1023/B:TRUC.0000021811.66966.1d

Moskal, B.M. (2003). Developing classroom performance assessments and scoring rubrics – Part 1. *ERIC Digest,* pp. 1-7

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, *7*(10), 23-31. https://doi.org/10.7275/q7rm-gg74

Muis, K. R., Ranellucci, J., Trevors, G., & Duffy, M. C. (2015). The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction*, *38*, 1-13. https://doi.org/10.1016/j.learninstruc.2015.02.001

Murphy, J. I. (2004). Using focused reflection and articulation to promote clinical reasoning: An evidence-based teaching strategy. *Nursing Education Perspectives*, *25*(5), 226-231.

Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, *2*(6), 165-178.

Neumann, J. A., & Forsyth, D. (2008). Teaching in the affective domain for institutional values. *The Journal of Continuing Education in Nursing*, *39*(6), 248-252. https://doi.org/10.3928/00220124-20080601-07

Newmann, F. M. (1996). Authentic assessment in social studies: Standards and examples. In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 359-380). Academic Press. https://doi.org/10.1016/B978-012554155-8/50014-4

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education*, *104*(4), 280-312. https://doi.org/10.1086/444136

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, *1*(2), 159. https://doi.org/10.1007/s10409-006-9595-6

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Merrill.

Oakleaf, M. J. (2006). *Assessing information literacy skills: A rubric approach* [Unpublished doctoral dissertation]. University of North Carolina at Chapel Hill, NC.

Oermann, M. H., Saewert, K. J., Charasika, M., & Yarbrough, S. S. (2009). Assessment and grading practices in schools of nursing: National survey findings Part I. *Nursing Education Perspectives*, *30*(5), 274-278.

Olfos, R., & Zulantay, H. (2007). Reliability and validity of authentic assessment in a web-based course. *Journal of Educational Technology & Society*, *10*(4), 156-173.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, *21*(3), 239-250. https://doi.org/10.1080/0260293960210304

Panadero, E., & Jönsson, 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

Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806-813. https://doi.org/10.1016/j.lindif.2012.04.007

Parmelee, D., Michaelsen, L. K., Cook, S., & Hudes, P. D. (2012). Team-based learning: A practical guide: AMEE Guide No. 65. *Medical Teacher*, *34*(5), e275-e287. https://doi.org/10.3109/0142159X.2012.651179

Payne, D.A. (2003). *Applied educational assessment* (2nd ed.). Wadsworth/Thomson Learning.

Peeters, M. J., Schmude, K. A., & Steinmiller, C. L. (2014). Inter-rater reliability and false confidence in precision: Using standard error of measurement within PharmD admissions essay rubric development. *Currents in Pharmacy Teaching and Learning*, *6*(2), 298-303. https://doi.org/10.1016/j.cptl.2013.11.014

Penny, J., Johnson, R. L., & Gordon, B. (2000). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing*, *7*(2), 143-164. https://doi.org/10.1016/S1075-2935(00)00012-X

Petkov, D., & Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. *Issues in Informing Science & Information Technology*, *3*, 499-511. https://doi.org/10.28945/910

Petraglia, J. (1998). *Reality by design: The rhetoric and technology of authenticity in education.* Routledge. https://doi.org/10.4324/9781410601254

Pindiprolu, S. S., Lignugaris/Kraft, B., Rule, S., Peterson, S., & Slocum, T. (2005). Scoring rubric for assessing students' performance on functional behavior assessment cases. *Teacher Education and Special Education*, *28*(2), 79-91. https://doi.org/10.1177/088840640502800201

Popham, W. J. (1997). What's wrong and what's right with rubrics. *Educational Leadership, 55*(2), 72-75. http://skidmore.edu/assessment/handbook/Popham_1997_Whats-Wrong_and-Whats-Right_With-Rubrics.pdf

Porcello, D., & Hsi, S. (2013). Crowdsourcing and curating online education resources. *Science*, *341*(6143), 240-241. https://doi.org/10.1126/science.1234722

Race, P., Brown, S., & Smith, B. (2004). *500 tips on assessment.* Routledge. https://doi.org/10.4324/9780203307359

Rafoth, M. A., Leal, L., & DeFabo, L. (1993). *Strategies for learning and remembering: Study skills across the curriculum.* National Education Association.

Rawekar, A., Garg, V., Jagzape, A., Despande, V., Tankhiwale, S., & Chalak, S. (2013). Team based learning: A controlled trial of active learning in large group settings. *Journal of Dental and Medical Sciences*, *7*(4), 42-8. https://doi.org/10.9790/0853-0744248

Resnick, L. B., Resnick, D. P., & DeStefano, L. (1993). *Cross-scorer and cross-method comparability and distribution of judgements of student math, reading, and writing performances: Results from the New Standards Project.* National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education, University of California, Los Angeles.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, *15*(1), 18–39. https://doi.org/10.1016/j.asw.2010.01.003

Rhodes, T. L., & Finley, A. P. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment.* Association of American Colleges and Universities.

Rogers, C. R. (1983). *Freedom to learn for the 80's.* Charles E. Merrill Publishing.

Rogers, G. M., & Sando, J. K. (1996). *Stepping ahead: An assessment plan development guide.* Rose-Hulman Institute of Technology.

Rosenzweig, P. (2014). *The halo effect and the eight other business delusions that deceive managers.* Simon and Schuster.

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31. https://doi.org/10.1207/s15326977ea1101_1

Schamber, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. *The Journal of General Education*, *55*(2), 103-137. https://doi.org/10.1353/jge.2006.0025

Schenck, A. D., & Daly, E. (2012). Building a better mousetrap: Replacing subjective writing rubrics with more empirically-sound alternatives for EFL learners. *Creative Education*, *3*(8), 1320-1325. https://doi.org/10.4236/ce.2012.38193

Schirmer, B. R., Bailey, J., & Fitzgerald, S. M. (1999). Using a writing assessment rubric for writing development of children who are deaf. *Exceptional Children*, *65*(3), 383-397. https://doi.org/10.1177/001440299906500308

Schmid, R. F., Miodrag, N., & Francesco, N. D. (2008). A human-computer partnership: The tutor/child/computer triangle promoting the acquisition of early literacy skills. *Journal of Research on Technology in Education*, *41(1*), 63-84. https://doi.org/10.1080/15391523.2008.10782523

Schmoker, M. (2006). *Results now: How we can achieve unprecedented improvements in teaching and learning.* Association for Supervision and Curriculum Development.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly*, *19*(2), 159-172. https://doi.org/10.1080/10573560308219

Shafer, W. D., Swanson, G., Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, *14*(2), 151-170. https://doi.org/10.1207/S15324818AME1402_3

Shipman, D., Roa, M., Hooten, J., & Wang, Z. J. (2012). Using the analytic rubric as an evaluation tool in nursing education: The positive and the negative. *Nurse Education Today*, *32(*3), 246-249. https://doi.org/10.1016/j.nedt.2011.04.007

Simon, M., & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research & Evaluation*, *7*(18), 103-121.

Slunt, K. M., & Giancarlo, L. C. (2004). Student-centered learning: A comparison of two different methods of instruction. *Journal of Chemical Education*, *81*(7), 985. https://doi.org/10.1021/ed081p985

Song, K. H. (2006). A conceptual model of assessing teaching performance and intellectual development of teacher candidates: A pilot study in the US. *Teaching in Higher Education*, *11*(2), 175-190. https://doi.org/10.1080/13562510500527701

Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2019). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 1-17. https://doi.org/10.1080/03075079.2019.1582015

Stein, B. S., Bransford, J. D., Franks, J. J., Vye, N. J., & Perfetto, G. A. (1982). Differences in judgments of learning difficulty. *Journal of Experimental Psychology: General*, *111*(4), 406. https://doi.org/10.1037/0096-3445.111.4.406

Stevens, D. D., & Levi, A. (2005). Leveling the field: Using rubrics to achieve greater equity in teaching and grading. *Essays on Teaching Excellence: Professional and Organizational Development Network in Higher Education*, *17*(1).

Stiggins, R. J. (2001). *Student-involved classroom assessment*. Prentice Hall.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, *12*(4), 437-475. https://doi.org/10.1023/A:1009084430926

Stuhlmann, J., Daniel, C., Dellinger, A., Kenton, R., & Powers, T. (1999). A generalizability study of the effects of training on teachers' abilities to rate children's writing using a rubric. *Reading Psychology*, *20*(2), 107-127. https://doi.org/10.1080/027027199278439

Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88. https://doi.org/10.1016/j.asw.2014.03.003

Supovitz, J. A., MacGowan, A., III, & Slattery, J. (1997). Assessing agreement: Interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment*, *4*(3), 237-259. https://doi.org/10.1207/s15326977ea0403_4

Svinicki, M. (2004). Authentic assessment: Testing in reality. *New Directions for Teaching and Learning*, *2004*(100), 23-29. https://doi.org/10.1002/tl.167

Swanson, H. L., O'Connor, J. E., & Cooney, J. B. (1990). An information processing analysis of expert and novice teachers' problem solving. *American Educational Research Journal*, *27*(3), 533-556. https://doi.org/10.3102/00028312027003533

Taggart, G. L., Phifer, S. J., Nixon, J. A., & Wood, M. (Eds.). (2001). *Rubrics: A handbook for construction and use*. Scarecrown Press

Tanner, D. E. (2001). Authentic assessment: A solution, or part of the problem? *The High School Journal*, *85*(1), 24-29. https://doi.org/10.1353/hsj.2001.0020

Taylor, J. M. (2011). *Interdisciplinary authentic assessment: cognitive expectations and student performance* [Unpublished doctoral dissertation]. Pepperdine University, CA.

Tedick, D. J. (2002). *Proficiency-oriented language instruction and assessment: Standards, philosophies, and considerations for assessment*. CARLA Working Paper Series. University of Minnesota, The Center for Advanced Research on Language Acquisition.

Tellez, K. (1996). Authentic assessment. In J. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 704-721). Macmillan.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, *9*(2), 1-10.

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, *36*(5), 509-547. https://doi.org/10.1080/02602930903540991

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). Mapping to know: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, *86*(2), 264-286. https://doi.org/10.1002/sce.10004

Tout, D., Pancini, G., & McCormack, R. (2014). Using mobile peer mentors for student engagement: Student rovers in the learning commons. *Higher Education Research & Development*, *33*(3), 595-609. https://doi.org/10.1080/07294360.2013.841645

Threats to reliability. (2012). http://dissertation.laerd.com/reliability-in-research-p2.php

Truemper, C. M. (2004). Using scoring rubrics to facilitate assessment and evaluation of graduate-level nursing students. *Journal of Nursing Education*, *43*(12), 562-564. https://doi.org/10.3928/01484834-20041201-11

Turley, E., & Gallagher, C. (2008). On the uses of rubrics: Reframing the great rubric debate. *English Journal*, 97(4), 87-92. http://0-www.jstor.org.source.unco.edu/stable/30047253

Van Merriënboer, J. J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Educational Technology.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist*, *41*(1), 19-31. https://doi.org/10.1207/s15326985ep4101_4

Vaughan, N. (2014). Student engagement and blended learning: Making the assessment connection. *Education Sciences*, *4*(4), 247-264. https://doi.org/10.3390/educsci4040247

Vendlinski, T., Underdahl, J., Simpson, E., & Stevens, R. (2002, June). Authentic assessment of student understanding in near-real time. *Proceedings of the National Educational Computing Conference, San Antonio, Texas.*

Visschers-Pleijers, A. J., Dolmans, D. H., De Leng, B. A., Wolfhagen, I. H., & Van Der Vleuten, C. P. (2006). Analysis of verbal interactions in tutorial groups: A process study. *Medical Education*, *40*(2), 129-137. https://doi.org/10.1111/j.1365-2929.2005.02368.x

Wallace, C. S., Prather, E. E., & Duncan, D. K. (2011). A study of general education astronomy students' understandings of cosmology. Part I. Development and validation of four conceptual cosmology surveys. *Astronomy Education Review*, *10*(1). https://doi.org/10.3847/AER2011029

Weigle, S. C. (2002). *Assessing writing.* Cambridge University Press. https://doi.org/10.1017/CBO9780511732997

Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, *16*(3), 194-209. https://doi.org/10.1016/j.jslw.2007.07.004

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice.* John Wiley & Sons.

Wellington, P., Thomas, I., Powell, I., & Clarke, B. (2002). Authentic assessment applied to engineering and business undergraduate consulting teams. *International Journal of Engineering Education*, *18*(2), 168-179.

Wiggins, G. P. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, *2(2)*, 1-3. https://doi.org/10.7275/ffb1-mm19

Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* Jossey-Bass.

Wiggins, G. P. (1998). *Educative assessment. Designing assessments to inform and improve student performance.* Jossey-Bass.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Metacognition in educational theory and practice* (pp. 277-304). Lawrence Erlbaum Associates Publishers.

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *Journal of Effective Teaching*, *7*(1), 3-14.

Wright, G. B. (2011). Student-centered learning in higher education. *International Journal of Teaching and Learning in Higher Education*, *23*(1), 92-97.

Wu, X. V., Heng, M. A., & Wang, W. (2015). Nursing students' experiences with the use of authentic assessment rubric and case approach in the clinical laboratories. *Nurse Education Today*, *35*(4), 549-555. https://doi.org/10.1016/j.nedt.2014.12.009

Young, C. (2013). Initiating self-assessment strategies in novice physiotherapy students: A method case study. *Assessment & Evaluation in Higher Education*, *38*(8), 998-1011. https://doi.org/10.1080/02602938.2013.771255

Young, L. P. (2009). Imagine creating rubrics that develop creativity. *English Journal*, *99*(2), 74.

The Role of Rubrics in Learning and Authentic Assessment

Young, M. (2007). *Bringing knowledge back in: From social constructivism to social realism in the sociology of education.* Routledge. https://doi.org/10.4324/9780203073667

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

BIOGRAPHIES

Clara Nkhoma is an Associate Lecturer in accounting at RMIT University Vietnam. Her research interests include student engagement, accountability, performance management and ethics education. Clara can be reached at clara.nkhoma@rmit.edu.vn

Professor Mathews Z. Nkhoma is a professor at RMIT University Vietnam. He holds a PhD in Information Security from University of East London, England. His major research topics are information systems security, information security investment model, ethical hacking, network defence, network security management and forensic computing. Mathews has international experience in teaching– leading Information systems and computer forensics courses in Africa, Europe, Middle East and Asia at both undergraduate and postgraduate levels. Mathews can be reached at mathews.nkhoma@rmit.edu.vn

Dr Susan Thomas is an Adjunct Associate Professor at RMIT University Vietnam and Professor at Okanagan School of Business, Canada. She has more than 20 years teaching experience and received several awards in teaching and learning, amongst which are Pro Vice Chancellor Award for Teaching, Pro Vice Chancellor Award for Administration, International Graduate Scholar Award, and Best Paper Presented in a Conference. She has published in top-tiered journal in the field of education, health care and management.

Nha Quoc Le joined RMIT University Vietnam in May 2018 in the role of Research Assistant and is responsible for providing research support to specific projects, coordinating the day-to-day research activities of the study as well. Before joining RMIT University Vietnam, he gained a Master's degree in Quantitative Finance from Christian-Albrechts-Universität of Kiel, Germany in September 2017. Currently, he places his focal point of study in education development and the efficiency of pedagogic supportive tools.