

The Validity and Instructional Value of a Rubric for Evaluating Online Course Quality: An Empirical Study

Ji Eun Lee and Mimi Recker
Utah State University

Min Yuan
University of Utah

Abstract

This study investigates the validity and instructional value of a rubric developed to evaluate the quality of online courses offered at a midsized public university. This rubric is adapted from an online course quality rubric widely used in higher education, the Quality Matters rubric. We first examine the reliability and preliminary construct validity of the rubric using quality ratings for 202 online courses and eliminate 12 problematic items. We then examine the instructional value of the rubric by investigating causal relationships between: (a) course quality scores, (b) online interactions between students, instructors, and content, and (c) student course performance (course passing rates). A path analysis model, using data from 121 online courses enrolling 5,240 students, show that only rubric items related to *learner engagement and interaction* have a significant and positive effect on online interactions, while only student-content interaction significantly and positively influence course passing rates.

Keywords: online course quality, rubric, online interactions, rubric reliability, rubric validity, quality matters rubric

Lee, J.E., Recker, M., & Yuan, M. (2020). The validity and instructional value of a rubric for evaluating online course quality: And empirical study. *Online Learning*, 24(1), 245-263. <https://doi.org/10.24059/olj.v24i1.1949>

The Validity and Instructional Value of a Rubric for Evaluating Online Course Quality: An Empirical Study

The number of college students taking online courses has increased dramatically over the past decade, with almost 31% of U.S. undergraduate students (about 5.2 million) having taken at least one course online as of the 2016 fall semester (McFarland et al., 2018). With this rapid growth in the number of online courses, evaluating their quality has taken on a new urgency. While many approaches have been developed to evaluate online course quality for example, surveys, checklists, observations, peer reviews, and expert reviews—one common way is through *quality rubrics* (Custard & Sumner, 2005; Jaggars & Xu, 2016; Roblyer & Wiencke, 2003; Yuan & Recker, 2019). With a quality rubric, a course can be rated along several constituent quality dimensions—for example, the Quality Matters (QM) rubric (Quality Matters, 2018) consists of eight dimensions,

such as learning objectives, instructional materials, learner support, accessibility, and usability, etc. Each of these dimensions may, in turn, be composed of one or more specific quality indicators (Custard & Sumner, 2005). In addition, for each indicator, rubrics often use rating scales and may be accompanied by a scoring guide.

While quality rubrics are commonly used in many higher education institutions, few rubrics have been empirically tested in terms of their reliability or validity (Yuan & Recker, 2015). Moreover, an often-ignored aspect of course quality is its influence on online interactions and student outcomes; in other words, the *instructional value* of the rubric. A key assumption is that a well-designed course following a proven instructional design theory will enhance student learning and engagement and thereby lead to improved outcomes (Reigeluth, 1999). Thus, a course that scores high on quality should result in better student outcomes than one receiving a low score. However, this relationship has seldom been examined in the literature (Jaggars & Xu, 2016).

The purpose of this article is twofold. The first is to test the validity of a rubric developed to evaluate the quality of online courses offered at a midsized public university. This rubric, called the AS rubric, was adapted from the QM rubric. The QM rubric is one of the most widely used rubrics in higher education and its design is informed by online learning research (Quality Matters, 2018). In particular, using the course quality scores from 202 online courses, we examined the preliminary construct validity of the AS rubric.

The second purpose is to examine the implicit logic linking online course quality to online interactions and student course performance. We investigated the causal relationships between course quality scores, online interactions between students, instructors, and content, and student performance as measured by their course passing rates. We characterized student and instructor online interactions in a subset of these online courses (the number of courses = 121; the number of students = 5,240) using the clickstream data automatically captured by the learning management system (LMS) for these courses. Finally, we examined the extent that the course quality measures, mediated by student and instructor interactions, influenced passing rates. The specific research questions guiding this research are:

1. To what extent is the AS online course quality rubric valid in measuring quality along a number of course quality dimensions? Which specific indicators are reliable (internal consistency reliability of the rubric) and valid (construct validity of the rubric)?
2. How do the course quality measures, when mediated by student and instructor online interactions, influence course passing rates?

Figure 1 articulates the logic underpinning this study: an online course that rates highly on quality along several key dimensions will positively influence the online interactions of its students and instructors and how they interact with content, which will ultimately lead to improved course performance. Figure 1 also illustrates how these three constructs are operationalized in our study.

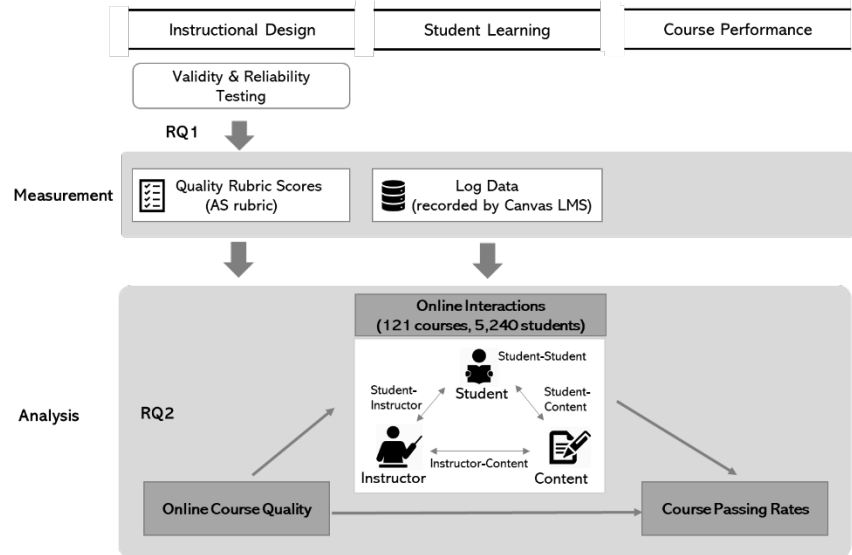


Figure 1. The study's logic linking instructional design to student course performance with measures for each component.

Review of Literature

In this section, we review the literature related to these three constructs shown in Figure 1. We first review the growing literature surrounding the use of course quality rubrics in higher education. We also specifically review the few studies that examine the relationship between online course quality scores and student learning outcomes. Finally, we describe a framework for characterizing and classifying interactions in online courses.

Course Quality Rubrics

We conducted a search of course quality rubrics in ERIC and Google Scholar with the following keywords: *online course*, *quality*, *rubric*, and *evaluation*. We also found rubrics from reviewing references of existing rubrics and getting recommendations from colleagues. These strategies yielded 31 rubrics. Ten course quality rubrics were ultimately selected based on the following criteria: they (a) were used for evaluating the quality of online courses; (b) consisted of more than two dimensions, with accompanying definitions of the dimensions; and (c) were used in higher education settings. Building on the approach used in a prior review of the quality rubric literature (Yuan & Recker, 2015), we examined online course quality rubrics used by higher education institutions in terms of three aspects: (a) *development process*, (b) *quality dimensions*, and (c) *and results of reliability and validity testing*.

First, in terms of the development process, most of the rubrics were adapted from other existing rubrics, rather than based on online learning theories or models (see Table 1). Regarding revisions to the rubrics, eight rubrics noted that they went through several rounds of revisions.

Table 1

Development Process, Reliability, and Validity of the Ten Rubrics Reviewed

No	Rubric	Development process	Reliability & Validity (publicly reported)
1	Checklist for Evaluating Online Courses (Southern Regional Education Board, 2006)	<ul style="list-style-type: none"> Developed based on Southern Regional Education Board’s standards for quality online courses 	<ul style="list-style-type: none"> Not reported
2	Quality Standards Inventory (Egerton & Posey, 2007)	<ul style="list-style-type: none"> Developed based on the principles of active learning and effective teaching 	<ul style="list-style-type: none"> Not reported
3	Online Course Design Rubric (New Mexico State University, 2011)	<ul style="list-style-type: none"> Developed based on QM Noted that “the rubrics are updated regularly.” 	<ul style="list-style-type: none"> Not reported
4	Online Course Best Practices Checklist (Palomar College, 2012)	<ul style="list-style-type: none"> Informed by a few existing rubrics (e.g., Blackboard, QM) Revised several times 	<ul style="list-style-type: none"> Reported that “a pilot test of the checklist was conducted” but specific results were not reported.
5	Quality Learning and Teaching Instrument (California State University 2015)	<ul style="list-style-type: none"> Informed by existing rubrics and models (e.g., QM, Community of Inquiry model) Revised several times 	<ul style="list-style-type: none"> Not reported
6	Online Educational Initiative Course Design Rubric (California Community College, 2016)	<ul style="list-style-type: none"> First version developed in 2014 by the OEI Development work group Revised based on feedback from instructors and reviewers 	<ul style="list-style-type: none"> Not reported
7	Exemplary Course Program Rubric (Blackboard Inc., 2017)	<ul style="list-style-type: none"> First developed in 2000 Reviewed and updated annually by Blackboard experts 	<ul style="list-style-type: none"> Not reported
8	Rubric for Evaluating Online Courses (University of North Dakota, 2017)	<ul style="list-style-type: none"> Developed based on a few existing rubrics (e.g., Blackboard) Revised several times 	<ul style="list-style-type: none"> Not reported
9	Quality Online Course Initiative Rubric (Illinois Center College, 2017)	<ul style="list-style-type: none"> Informed by existing rubrics Brainstormed dimensions first and then chunked into categories Revised several times 	<ul style="list-style-type: none"> Not reported
10	Quality Matters (QM): Course Design Rubric Standards (2018)	<ul style="list-style-type: none"> Informed by a few research articles, and revised based on users’ inputs Revised for a few versions 	<ul style="list-style-type: none"> Improvement process reported (Shattuck et al., 2014) Measured “rater agreement.”

Second, with regard to quality dimensions, although each rubric used slightly different terms, our review found five common dimensions for measuring online course quality across the rubrics. These were: (a) course design and introduction, (b) learning objectives and assessment, (c) interaction and collaboration, (d) learning resources and support, and (e) course technology and accessibility. However, the rubrics also showed differences in their evaluation focus. For instance, Rubric #10 (Quality Matters, 2018) consisted of 42 weighted items with almost 30% of the weight

addressing “learning objectives and assessment” and only 11% of the weight focused on “interaction and collaboration.” In contrast, Rubric #6 (California Community College, 2016) emphasized “course technology and accessibility” with 48% of the total items related to these issues.

Finally, rubrics require sufficient levels of *reliability* and *validity* (Roblyer & Wiencke, 2003). Despite the importance of establishing reliability and validity of rubrics, none of the reviewed rubrics publicly reported the results of reliability or construct validity tests. Only two rubrics (Rubric #4 and #10 in Table 1) noted that they underwent empirical testing, such as a measurement of rater agreement, but details were not reported. This lack of reliability or validity testing calls into question the rubrics’ overall suitability for rigorously evaluating online course quality (Yuan & Recker, 2015).

To summarize, the ten rubrics reviewed in this study show similarities in the dimensions addressed and the rating scales used, but they differed in their focus for evaluation. These differences seem reasonable, as all higher education institutions have different needs, interests, and criteria for evaluating online courses (Britto, Ford, & Wise, 2013). However, from a research perspective, key questions remain: which dimensions are more important in evaluating the quality of an online course? Which dimensions better predict student performance?

Course Quality and Student Learning Outcomes

Our literature review suggests that rubrics for measuring course quality have been validated mostly in terms of the opinions and perceptions of faculty and students, rather than in terms of construct validity or relationships to learning outcomes (Hixon, Barczyk, Ralston-Berg, & Buckenmeyer, 2016). Empirical studies (Jaggars & Xu, 2016; Lee, 2014; Liu et al., 2010; Sun et al., 2008; Swan et al., 2012) have found that a course with high quality scores measured by rubrics resulted in higher student learning outcomes in terms of course performance or satisfaction than one receiving low quality scores. However, studies also showed that not all scores on dimensions of the rubrics significantly predicted learning outcomes (Jaggars & Xu, 2016; Lee, 2014; Sun et al., 2008). For instance, Jaggars & Xu (2016) explored the relationship between rubric scores from 23 online courses and student final grades at two community colleges in the U.S. Results revealed that among the four rubric dimensions, only the “interpersonal interaction” dimension had a statistically significant and positive impact on student final grades. Thus, while well-organized courses or well-described learning objectives might be desirable, these quality aspects may not lead to better learning outcomes per se.

Characterizing Interactions in Online Learning

Interactions among learners, instructors, and content are integral components of online education (Bernard et al., 2009). A widely used framework for examining interactions in online education is Moore’s (1989) *interaction framework*. This framework classifies interactions into three types: Student-Instructor, Student-Student, and Student-Content.

Later, Anderson and Garrison (1998) expanded Moore’s framework by differentiating between Student-Content and Instructor-Content interaction. These four types of interactions are defined by Anderson (2008) as Student-Instructor (SI), Student-Student (SS), Student-Content (SC), and Instructor-Content (IC). SI interaction refers to communication between learners and experts, which includes instructor feedback, support, and encouragement to learners. SS interaction is defined as communication between one learner and other learners, including

collaborative or cooperative settings. SC interaction includes student activities such as reading course materials, watching lecture videos, and completing assignments. IC interaction refers to instructors creating, monitoring, or modifying content or learning activities.

Many empirical studies have examined how the strength of interactions is associated with student learning outcomes, such as their performance or satisfaction (Borokhovski et al., 2012; Choi, Lee, Hong, Lee, Recker, & Walker, 2016; Hoey, 2017; Ke, 2013; Kuo et al., 2013; Murray et al., 2012; Sher, 2009). However, the effects of each interaction type on learning outcomes have not been found to be equal. Our review found that studies yielded different results depending on the outcome variable studied.

First, studies that used measures of student course performance as dependent variables indicated that the effects of SC or SS interaction were larger than the effect of SI interaction on student performance. For instance, Bernard et al. (2009) reviewed 74 empirical studies to examine the effects of three types of interaction (SS, SI, SC) strength on student performance. The results of a meta-analysis revealed that the effects of SS and SC interactions were significantly larger than the effect of SI interaction on performance. Similarly, in other studies, SS or SC interactions (Ke, 2013), SS interaction (Borokhovski et al., 2012; Choi et al., 2016), or SC interaction (Murray et al., 2012) had significant and positive influences on student performance.

Second, studies that used student affective outcomes as dependent variables tended to show somewhat different results. For instance, in the meta-analysis by Bernard et al. (2009), the effect of SS interaction was significantly larger than the effects of SC or SI interactions on student attitudes. However, a study by Kuo et al. (2013) produced opposite results, finding that SC and SI interactions were significant predictors of student satisfaction, while SS interaction was not. To summarize, our review found that the effects of each interaction type differed depending on the dependent variable used in the study and the characteristics of interactions analyzed.

Methods

Course Quality Rubric

This study used course quality rating scores collected through a rubric used at a mid-sized public university in the U.S. The rubric was developed collaboratively by instructional designers at an Academic Support (AS) unit in order to support instructional designers in better designing online courses as well as ensuring online course quality at this university. The AS rubric was adapted from the well-established and reliable QM rubric and consists of nine dimensions (course organization, course introduction and syllabus, learning objectives, assessments and activities, resources and materials, interaction and learner engagement, accessibility, course technology, and learner support) and 51 items to measure online course quality.

However, we identified several problems with these predefined dimensions. First, the number of items measuring each quality dimension, which influences the coefficients of internal consistency and reliability (Drost, 2011), varied widely across the dimensions (from 3 to 12 items). Second, some items did not adequately reflect their dimension, which raises content validity issue. For instance, one item in the “course instruction and syllabus” dimension, “provides clear expectations for student response, engagement, and participation,” also aligned to the “interaction and learner engagement” dimension. For these reasons, we decided to ignore the predefined

dimensions and generate new ones using the results of an exploratory factor analysis, described below.

Research Context and Participants

To measure the preliminary construct validity of the AS rubric (RQ1), we used course quality scores collected from the ratings of 202 online courses offered at this university from 2012 to 2016. Among the 2,797 courses offered during this period, the instructional designers randomly selected 202 courses and evaluated their course quality using the AS rubric. The courses included both undergraduate (173 courses, 85.6% of the sample) and graduate level courses (29 courses, 14.4% of the sample) from various academic disciplines. Each course was rated by one instructional designer in the AS unit at the beginning of the semester. The items were rated on a two-point scale (Yes = 1, No = 0). Note that no responses were coded as null.

To measure the level of online interactions in each course (RQ2), we categorized instructor and student clickstream data automatically collected by the university’s LMS into the four types of interactions as defined by the framework described above (see Table 2). Of the original sample of 202 courses, 81 lacked LMS interaction data or student final grades and were excluded from further analysis. The remaining 121 courses enrolled a total of 5,240 students. All measures were converted to Z-scores before computing the average level of interaction. We also measured student course performance in terms of passing rates. This was computed by dividing the number of students who successfully passed the courses (receiving grades of A, B, C, or D) by the number of students enrolled in each course. Among these students, 169 students (3%) received a grade of W (Withdrawal), indicating that the students dropped the course after the first three weeks of the semester.

Table 2

Summary of LMS Variables Used to Measure the Four Types of Interaction

Online Interactions	LMS Variables	Measures
Instructor-Content (IC)	ic_atta # of attachments posted by an instructor	$\frac{ic_{atta} + ic_{disc} + ic_{wiki} + ic_{quiz} + ic_{assi}}{5}$
	ic_disc # of discussion topics posted by an instructor	
	ic_wiki # of wiki topics posted by an instructor	
	ic_quiz # of quizzes posted by an instructor	
	ic_assi # of assignments posted by an instructor	
Student-Content (SC)	sc_atta Avg. # of attachments viewed by a student	$\frac{sc_{atta} + sc_{disc} + sc_{wiki} + sc_{quiz} + sc_{assi}}{5}$
	sc_disc Avg. # of discussions viewed by a student	
	sc_wiki Avg. # of wiki topics viewed by a student	
	sc_quiz Avg. ratio of quizzes completed by a student	
	sc_assi Avg. ratio of assignments completed by a student	
Student-Student (SS)	ss_disc Avg. # of discussion messages (initial messages and replies) posted by a student	- ss_disc
Student-Instructor (SI)	si_disc # of discussion messages (initial messages and replies) posted by an instructor	- si_disc

Note. The course is the unit of analysis. All interaction measures were converted to Z-scores.

Data Analysis

Before examining the validity of the rubric (RQ1), the internal consistency reliability of the AS rubric was measured using Kuder-Richardson formula-20 (KR-20) with two-point measurement data. Specifically, we used a stepwise procedure to find unreliable items and to maximize scale reliability (Raubenheimer, 2004). In the stepwise procedure, the least reliable item is removed, as indicated by the expected increase in KR-20 coefficient for the subscale. Then, the next least reliable item is removed, and the analysis is repeated until the removal of items does not lead to an increase in reliability.

To examine the preliminary construct validity of the rubric, we conducted an exploratory factor analysis (EFA) as we had little theoretical or empirical basis for the rubric’s design. Since our data are dichotomous, we computed *tetrachoric correlation coefficients* and then conducted an EFA using these coefficients. For the extraction factor rotation methods, we chose unweighted least-squares (ULS) extraction with Promax rotation, the recommended method for the analysis of tetrachoric correlation coefficients (Han et al., 2001).

For RQ2, we conducted a path analysis to investigate the relationships between online course quality scores, online interactions, and passing rates. The path model tested three hypotheses: (a) the online course quality scores influence all variables (the four types of interactions) and passing rates; (b) the four types of interactions influence passing rates, and; (c) the online interactions mediate the influence of online course quality scores on passing rates. R Studio with the *psych* and *lavaan* packages was used for all analyses.

Results

Research Question 1: Reliability and the Preliminary Construct Validity of the AS Rubric

The first research question examined the reliability and the validity of the AS quality rubric using its quality dimensions and items. To answer this question, we conducted an internal consistency reliability analysis and an EFA. The initial KR-20 coefficient for 51 items was .82. Next, the stepwise procedure was performed to maximize reliability. As a result, eight items were eliminated (16% of the total) (see Table 3), and the KR-20 coefficient for 43 items increased to .87. As summarized in Table 3, four of the eliminated items (item #39, #40, #41, #42) were related to the “accessibility” dimension. The other four eliminated items (item #28, #30, #31, #47) related to course technology issues

Table 3
The Items Eliminated from the Reliability Test and the EFA

	Item no.	Descriptions
Items removed from the reliability test	item40	Scanned PDF documents are made screen readable with OCR technology.
	item41	Images used for learning have a visual description.
	item39	Audio is captioned or transcribed.
	item47	Course provides sufficient instructions for students on use of tools and media.
	item31	No unreasonable software requirements.
	item42	Images have an alt tag.
	item30	Resources & materials can be accessed with multiple operating systems.
	item28	Resources & materials are easily accessed and used.

Items removed from the EFA	item11	Provides clear expectations for instructor response and engagement.
	item08	Evaluation methods and assessment activities are clearly outlined.
	item29	Purpose of each element is explained
	item32	Learner engagement and interaction activities promote achievement of learning objectives.

Next, we conducted an EFA using the remaining 43 items to examine the preliminary construct validity of the rubric. The results of Bartlett’s test of sphericity ($\chi^2[903] = 16200.13, p < .05$) and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (KMO = .70) indicated that our data were suitable for performing a factor analysis (Yong & Pearce, 2013). Forty-three items were analyzed using an ULS extraction method with Promax rotation. For the convergent validity, we used cut-off loadings of 0.4. Next, to determine the number of factors to retain for rotation, we checked eigenvalues (Kaiser’s rule) and performed a parallel analysis. The results indicated that the nine-factor solution had the cleanest structure (i.e., fewest cross-loadings and no factors with fewer than three items).

Table 4 shows the results of factor loadings for the 43 items. The nine-factor solution explained 73% of the total variance. Among the 43 items, another four items were eliminated because one cross-loaded onto two factors, and the other three did not have primary factor loadings of .4 or above. These four items tended to have imprecise descriptions or criteria to evaluate course quality, perhaps making use by raters difficult (see Table 3).

Table 4
Results of Factor Loadings for AS Rubric Items (43 items)

Items	1	2	3	4	5	6	7	8	9
Eigenvalues	5.99	4.08	3.70	4.03	3.09	3.01	2.51	2.59	2.44
% of variance	0.14	0.09	0.09	0.09	0.07	0.07	0.06	0.06	0.06
Cumulative %	0.14	0.23	0.32	0.41	0.49	0.56	0.61	0.67	0.73
item19	0.59	-0.13	-0.25	-0.20	0.26	-0.03	0.09	0.12	0.37
item22	0.42	0.13	0.21	0.08	0.22	0.16	-0.28	-0.04	0.29
item23	0.87	-0.15	0.15	-0.12	0.01	0.20	0.07	0.12	-0.43
item24	0.44	-0.09	-0.02	0.19	0.06	0.34	-0.17	0.24	-0.18
item25	0.58	0.32	-0.09	-0.23	0.15	0.00	0.29	-0.05	0.14
item26	0.68	0.18	-0.12	0.28	-0.17	-0.02	0.12	-0.13	0.16
item27	0.83	-0.01	0.11	0.16	-0.25	0.09	0.01	-0.01	0.03
item43	0.64	-0.15	0.09	0.20	0.20	0.06	-0.05	0.33	-0.10
item44	0.82	0.31	-0.14	0.07	-0.03	-0.01	-0.03	-0.30	-0.03
item48	0.51	0.13	0.09	-0.09	-0.12	-0.16	0.06	0.28	-0.01
item01	0.02	0.89	-0.22	-0.04	-0.01	-0.01	-0.06	-0.01	0.11
item02	0.08	0.71	0.06	-0.08	-0.13	0.16	-0.14	0.11	0.04
item03	0.17	0.84	0.14	0.10	-0.21	0.16	-0.22	-0.07	-0.09
item04	0.12	0.62	0.18	0.20	0.10	-0.01	0.03	-0.06	-0.02
item49	0.12	-0.03	0.78	-0.13	0.00	0.10	0.10	0.02	0.00
item50	-0.03	0.06	1.02	0.04	0.00	0.04	-0.05	-0.13	-0.07
item51	-0.08	0.06	0.98	0.06	0.02	0.01	-0.05	-0.12	0.00
item13	0.18	0.10	0.03	0.76	0.09	-0.06	0.38	-0.04	-0.13
item14	-0.06	0.30	0.20	0.55	0.22	0.07	-0.05	0.12	-0.08
item34	0.03	-0.05	0.16	0.73	0.21	-0.13	0.09	-0.37	0.18
item35	0.06	-0.15	-0.18	0.74	-0.22	0.35	0.24	-0.12	0.30
item36	0.39	0.03	-0.18	0.54	0.07	0.14	0.01	-0.06	0.03
item37	-0.01	0.14	-0.26	0.47	0.08	0.03	-0.11	0.27	0.07

Items	1	2	3	4	5	6	7	8	9
item16	-0.15	0.00	0.00	0.24	0.98	0.10	0.04	-0.11	-0.30
item17	-0.08	-0.10	-0.03	0.11	0.81	-0.19	-0.01	-0.05	0.13
item18	0.16	-0.17	0.03	-0.03	0.78	-0.06	0.06	0.13	0.11
item05	-0.05	0.26	0.12	0.03	-0.20	0.59	0.12	0.00	-0.07
item06	0.17	0.10	-0.02	0.02	0.08	0.76	0.07	0.08	-0.13
item12	0.12	-0.44	0.33	-0.07	-0.15	0.46	0.39	0.03	0.21
item33	0.26	-0.03	0.19	0.27	-0.02	0.44	-0.09	0.09	0.21
item07	-0.35	0.25	0.08	-0.04	0.26	0.11	0.41	0.28	0.19
item09	0.17	0.06	0.15	-0.11	0.10	0.34	0.49	-0.09	0.09
item10	0.09	-0.34	-0.08	0.40	-0.04	0.03	0.99	-0.01	0.08
item38	-0.24	0.02	0.02	0.05	0.00	0.33	0.00	0.42	0.39
item45	0.08	0.17	-0.28	-0.25	-0.11	0.22	0.08	0.96	0.11
item46	0.40	-0.19	0.17	-0.03	0.10	-0.19	-0.08	0.66	0.10
item15	-0.23	0.32	0.14	0.00	0.23	0.37	-0.12	-0.14	0.54
item20	-0.02	0.13	0.32	0.14	-0.28	-0.37	-0.09	0.18	0.44
item21	0.08	-0.10	-0.13	0.16	-0.06	-0.16	0.28	0.17	0.67
item11*	-0.24	0.29	0.01	0.46	-0.08	0.21	0.43	0.13	-0.09
item08**	0.08	0.11	0.34	0.00	0.13	-0.49	0.36	0.07	0.09
item29**	0.39	0.35	-0.13	0.02	0.10	-0.03	0.06	0.19	-0.01
item32**	0.24	0.23	-0.24	0.37	-0.01	0.06	-0.07	0.17	0.24

Note. Factor loadings < .4 are suppressed. * Item cross-loaded onto multiple factors. ** Items without primary factor loadings of .4 or above.

Finally, Table 5 summarizes the nine factors, their labels, and their 39 items based on the EFA. Factor 1 accounted for the highest amount of the total variance (14%) among the nine factors. Ten items displayed meaningful loadings (greater than .40) for this factor and all the items related to student activities or course content. This factor was labeled “Learning Activities & Materials.”

Table 5
Summary of New Factors and Their Items Based on the EFA

EFA constructs and labels	Items from AS rubric
Factor 1 (Learning Activities & Materials)	item19 Assessments and activities are consistent with the course objectives and resources.
	item22 Activities provide students with opportunities to receive feedback early and frequently, specifically in preparation for high stakes assessments.
	item23 Course includes assessments and activities that are problem-centered or application-oriented in nature.
	item24 Students are encouraged to integrate new concepts into regular practice and understanding through demonstration, reflection, creation, or similar activities.
	item25 Resources & materials support learning objectives.
	item26 Resources & materials are sufficient for students to learn the subject.
	item27 Resources, materials, and instructor interactions activate students’ prior learning and experiences while introducing new concepts.
	item43 Tools and media support the learning objectives.
	item44 Tools and media are appropriately chosen and appropriately varied to enhance student interactivity with course content.
	item48 Course provides additional tutorials/resources as needed to accomplish objectives.

The Validity and Instructional Value of a Rubric for Evaluating Online Course Quality: An Empirical Study

EFA constructs and labels	Items from AS rubric	
Factor 2 (Course Introduction & Design)	item01	Upon first entering the course, students can easily find the course syllabus and introductory materials.
	item02	The progression of course content and activities is easy to find, clearly outlined, and appropriately segmented into units or modules.
	item03	Course appears visually clean, consistent, and appealing on the home page and throughout.
	item04	A course introduction orients student to the course environment and suggests the relevance of course materials and activities to students and/or program goals.
Factor 3 (Learner Support)	item49	Course provides technical support services link/description.
	item50	Course provides academic support services link/description.
	item51	Course provides student support link/description.
Factor 4 (Learner Engagement & Interaction)	item13	Provides clear expectations for student response, engagement, and participation.
	item14	Provides clear expectations for student etiquette in participation.
	item34	A means for making course announcements is clearly available and used regularly to encourage student completion and participation and to connect course content with current events and research.
	item35	Course design fosters interaction with other students.
	item36	Course design fosters interaction with content.
	item37	Appropriate synchronous or asynchronous means are provided for students to ask questions and receive answers from the instructor and/or students.
	Factor 5 (Learning Objectives)	item16
	item17	Objectives are measurable.
	item18	Objectives are consistent with the course material/assessments/assignments.
Factor 6 (Course Facilitation)	item05	Course has an instructor introduction.
	item06	Students have an opportunity to introduce themselves.
	item12*	Course fees, if any, are explained.
	item33	Course design fosters interaction with instructors.
Factor 7 (Course Information)	item07	The course grading policy is clearly stated.
	item09	Course technology requirements are addressed up front, if applicable.
	item10	Textbook information and other materials requirements are provided.
Factor 8 (Course Technology)	item38	Course has a statement directing students with ADA-documented disability to the DRC for reasonable accommodations as needed.
	item45	Tools and media are as easy to use as is reasonably possible.
	item46	Tools and media are sufficiently compatible with web and other applicable standards.
Factor 9 (Course Management)	item15	Syllabus addresses course-appropriate policies, including academic honesty, harassment, withdrawal and I-grades, and the student grievance process.
	item20	Appropriate pacing mechanisms (due dates, reminders, follow-ups) are used to ensure timely student completion and regular engagement.
	item21	Specific descriptive criteria are provided for the evaluation of student's work and participation, ideally in the form of a rubric.

Note. * Item does not fit well in category

Factors 2, 3, and 4 each explained 9% of the total variance. The four items loading onto Factor 2 related to aesthetic dimensions of the course or its introductory materials. This factor was labeled “Course Introduction & Design.” The three items loading onto Factor 3 dealt with whether academic or technical support links/descriptions are provided in the courses (labeled “Learner Support”). Six items displayed significant loadings for Factor 4 related to interaction, student participation, and engagement in courses (labeled “Learner Engagement & Interaction”).

Factors 5 and 6 each explained 7% of the variance. Factor 5 consisted of three items and was labeled “Learning Objectives.” Four items displayed meaningful loadings for Factor 6. Three items (item5, item6, item33) dealt with facilitating the courses (labeled “Course Facilitation”). However, one item (item12: “Course fees, if any, are explained”) did not seem to measure the same construct as other items, which implies that revisions to the rubric are needed.

Factors 7, 8, and 9 each explained 6% of the total variance. The three items loaded onto Factor 7 dealt with course policy or requirements (labeled “Course Information”). Factor 8 consisted of three items related to course technology issues (labeled “Course Technology”). The three items showing meaningful loadings for Factor 9 dealt with course management issues such as syllabus, pacing mechanism, and evaluation of student work (labeled “Course Management”).

Research Question 2: Instructional Value of the Rubric

The second research question investigated how course quality measures, when mediated by student and instructor online interactions, influenced course passing rates. We used a path analysis to model the influence of course quality scores on the four types of online interactions and passing rates. Table 6 summarizes the descriptive statistics for course quality rubric scores, online interactions, and passing rates. For course quality scores, we computed average rubric scores for the nine factors identified by the EFA.

Table 6
Descriptive Statistics of all Variables Included in the Path Model (N = 121 courses, 5,240 students)

Variables		M	SD	Min.	Max.
Course quality scores (rubric scores)	Factor 1: Learning Activities & Materials	0.92	0.16	0.10	1.00
	Factor 2: Course Introduction & Design	0.87	0.26	0.00	1.00
	Factor 3: Learner Support	0.87	0.30	0.00	1.00
	Factor 4: Learner Engagement & Interaction	0.76	0.30	0.00	1.00
	Factor 5: Learning Objectives	0.88	0.27	0.00	1.00
	Factor 6: Course Facilitation	0.83	0.25	0.00	1.00
	Factor 7: Course Information	0.95	0.17	0.00	1.00
	Factor 8: Course Technology	0.93	0.17	0.00	1.00
	Factor 9: Course Management	0.80	0.26	0.00	1.00
Online interactions (recorded by LMS)	Instructor-Content interaction*	0.00	0.67	-0.95	3.37
	Student-Content interaction*	0.00	0.58	-1.22	2.44
	Student-Student interaction*	0.00	0.99	-0.63	4.97
	Student-Instructor interaction*	0.00	0.99	-0.73	4.86
Course passing rate (ratio)		0.90	0.12	0.45	1.00

Note. The course quality scores are binary. * All interaction measures were converted to Z-scores.

First, we performed a path analysis using the initial model, with the direct effect of the course quality scores on course passing rates represented as path c, the direct effect of online interactions on course passing rates represented as path b, and the indirect effect of course quality scores on course passing rates represented as path a (see Figure 2). The model was statistically significant ($\chi^2[6] = 89.34; p < .05$), but it did not have a satisfactory model fit (Comparative Fit Index [CFI] = .37, recommended to be greater than .90) and included nonsignificant paths.

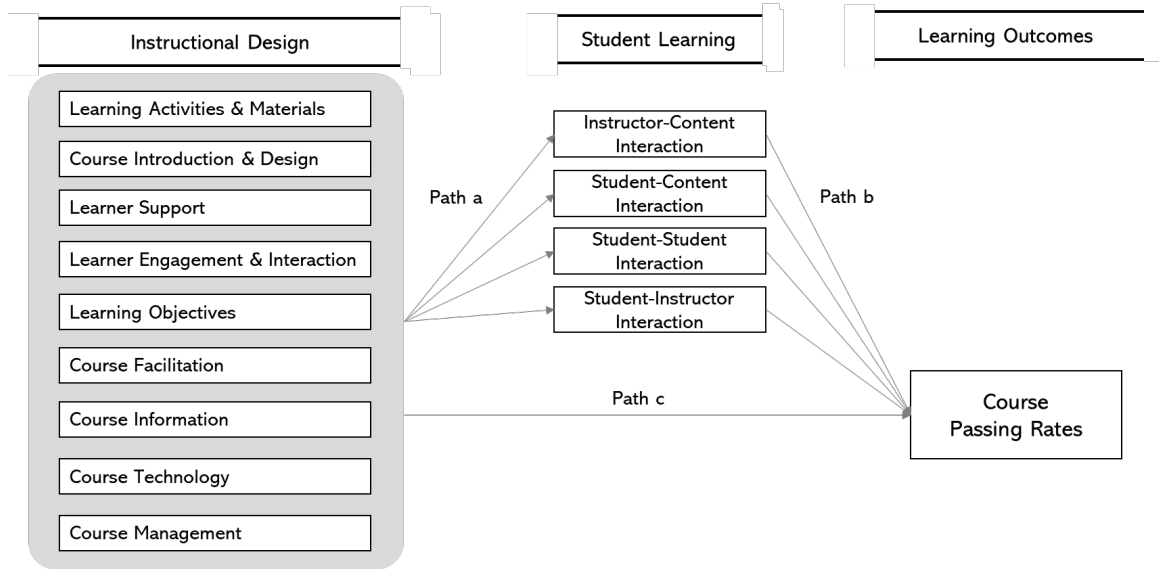


Figure 2. Path diagram for the initial model of the relationships among the course quality scores, online interactions, and course passing rates. (Note: Path a is from each of the nine factors to the interaction variables.)

We therefore dropped the nonsignificant paths and reconducted the path analysis, which showed good model fit ($\chi^2[6] = 14.26; p < .05$, CFI = .91, RMSEA = .11). Figure 3 shows the results with the standardized regression coefficients. In the revised model, all path coefficients were significant at the .05 level except for one path (Course Facilitation - Passing rate, $\beta = .155, p > .05$).

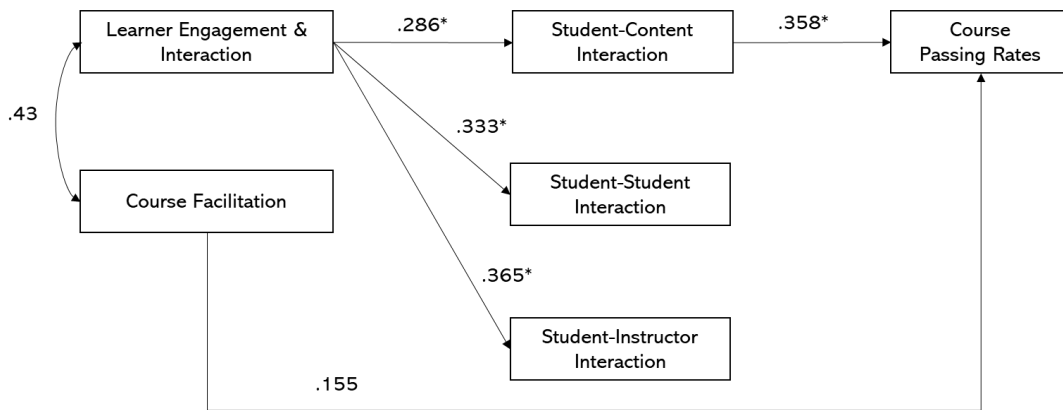


Figure 3. Path diagram for the final model.

Regarding the causal relationships between online course quality scores and online interactions, “learner engagement & interaction” scores had significant influences on Student-Content ($\beta = .286, p < .05$), Student-Student ($\beta = .333, p < .05$), and Student-Instructor interactions ($\beta = .365, p < .05$). Finally, Student-Content interaction had a significant direct effect on passing rate ($\beta = .358, p < .05$). The R-squared value indicates that approximately 16.3% of the variance in passing rate is explained by this model.

Discussion

This study examined the preliminary construct validity and instructional value of an online course quality rubric, the AS rubric. Instructional value was investigated in terms of the relationships between course quality, as measured by the AS rubric scores, online interactions between students, instructors, and content as automatically captured by the Canvas LMS, and student course passing rates.

For RQ1, the internal consistency reliability test for the AS quality rubric revealed eight unreliable items. Four were related to course accessibility, while the other four were related to course technology or course materials and resources. In addition, we found that some of the removed items did not use precise terms or clear guidelines in terms of evaluating course quality. For instance, the item “no unreasonable software requirements” did not define “unreasonable.” Similarly, in the case of the item “course provides sufficient instructions for students on use of tools and media,” the criteria for “sufficient” can be subjectively interpreted. Internal consistency reliability can be improved by using precise terms, clear guidelines, and making instructions as explicit as possible (Cohen et al., 2007). The EFA revealed four additional problematic items that either loaded on multiple factors or did not significantly load on any factor. The EFA identified nine factors, explaining 73% of the total variance. Among these nine factors, “learning activities & materials” explained the highest amount of total variance in course quality.

For RQ2, we modeled the causal relationships between the online course quality scores, the four types of online interactions captured by the LMS, and passing rates using a path analysis. First, results show that only rubric scores related to the “learner engagement and interaction” construct had a positive and significant effect on online interactions. The quality scores of “learner engagement and interaction” had the largest effect on SI interaction, followed by SS and SC interactions. Thus, online courses that are designed to encourage student participation and interaction with other students appear to not only have a higher level of SS interaction but also a higher level of SC and SI interactions. The quality measures for the other dimensions did not have a significant impact on any of the types of online interactions. While these dimensions address course features that are certainly desirable aspects to include in course design, they may not contribute to enhanced online interactions per se.

Second, in terms of the associations between the four types of interactions and passing rates, only SC interaction had a significant and positive effect on passing rates. This aligns with previous findings that SC interaction positively influenced performance (Bernard et al., 2009; Ke, 2013; Murray et al., 2012). We also note that SS interaction did not have a significant effect on passing rates. One reason for this result might be contextual differences as this study included courses from various academic disciplines. Indeed, one study (Ke, 2013) found that there were significant differences between disciplines in terms of the amount and type of online interactions.

Lastly, in terms of the relationship between the course quality scores and passing rates, the scores for one construct, “course facilitation,” had positive and significant influences on passing rates in the initial model, but not in the final model. However, scores on the “learner engagement and interaction” construct had a positive and significant effect on SC interaction, which, in turn, significantly and positively influenced passing rates. Thus, the results imply that course design elements related to “learner engagement and interaction” are an important aspect of course quality, indirectly contributing to course performance. Another study (Jaggars & Xu, 2016) reported a similar result in that the “interpersonal interaction” dimension of a quality rubric had a significant and positive impact on student final grades, while other dimensions of the rubric did not. In addition, while the final path model explained only 16.3% of the variability in passing rates, it is important to note that many other factors, in particular, student-related factors (e.g., academic background, relevant experiences), also influence successful course completion (Lee & Choi, 2011).

Limitations and Future Research

Several limitations to this research are important to note. In terms of the AS rubric, although the quality of over 200 online courses was measured, all came from a single university with its own institutional culture. Also, the rubric was only applied by one rater thus making it impossible to determine another important form of reliability, inter-rater reliability. Finally, the rubric used a binary score while a Likert scale may have increased the usability of the rubric (Yuan & Recker, 2015). In addition, our data were also drawn from various academic disciplines. As previously mentioned, one study (Ke, 2013) found significant disciplinary differences in online interaction patterns. Therefore, future research should consider the quality of online interactions using a disciplinary lens. Future work should also consider how results from this study inform rubric design to improve validity and instructional value. Finally, future work should examine the influence of course design and interaction variables on other important kinds of student learning outcomes (e.g., satisfaction, perseverance).

Conclusions

While the AS rubric was based on the widely used and reliable QM rubric, almost one-fourth of the rubric items were identified as problematic. This concerning result has implications for other quality rubrics used in higher education institutions because: (a) most of the rubrics reviewed in the literature were adapted from existing rubrics, rather than based on empirical testing or online learning models and (b) none of the rubrics reported results from reliability or validity tests. In particular, a lack of construct validity may result in misinterpretations of a construct, as well as raise doubts about the suitability and credibility of the measurement tool (Cohen et al., 2007; Yuan & Recker, 2015). Thus, more empirical studies are needed to establish the reliability and validity of existing course quality rubrics.

From a practical perspective, this study has several implications. During the course design stage, instructors and course designers could consider adding different strategies to promote students’ engagement and interactions, for example by using games and simulations, providing hands-on activities, and building an online course community using social networks. During the course review process, course designers could consider providing rubric definitions and guidelines, especially for items that are more subjective. They could also consider revising items related to course accessibility and technology use to make them easier to apply.

At the university level, although different higher education institutions might have different needs and criteria for evaluating online courses, a quality rubric plays an important role in identifying and addressing elements deemed important to instructional design (e.g., accessibility, course objectives). It is important to consider to what extent these elements serve to influence (or not) subsequent online interactions and learning outcomes. Many factors, stakeholders, and decisions influence the design of online courses and these results are revealing in terms of identifying those that seem to have a greater impact on students and providing guides for instructors and instructional designers on their course design process.

References

- Anderson, T. (2008). *The theory and practice of online learning*. Athabasca University Press.
- Anderson, T. D., & Garrison, D. R. (1998). Learning in a networked world: New roles and responsibilities. In C. C. Gibson (Ed.), *Distance learners in higher education* (pp. 97–112). Atwood Publishing.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243–1289.
- Blackboard Inc. (2017). *Blackboard exemplary course program rubric*. Blackboard. Retrieved from <https://www.blackboard.com/resources/are-your-courses-exemplary>
- Borokhovski, E., Tamim, R., Bernard, R. M., Abrami, P. C., & Sokolovskaya, A. (2012). Are contextual and designed student-student interaction treatments equally effective in distance education? *Distance Education, 33*(3), 311–329.
- Britto, M., Ford, C., & Wise, J. M. (2013). Three institutions, three approaches, one goal: Addressing quality assurance in online learning. *Online Learning Journal, 17*(4).
- California Community College. (2016). *Course design rubric for the online education initiative*. Retrieved from http://ccconlined.org/wp-content/uploads/2015/11/OEI_Rubric_Edited-ACC.pdf
- California State University (2015). *Quality learning and teaching*. Retrieved from <https://www.csun.edu/it/qlt>
- Choi, H., Lee, J. E., Hong, W. J., Lee, K., Recker, M., & Walker, A. (2016). Exploring learning management system interaction data: Combining data-driven and theory-driven approaches. In T. Barnes, M. Chi & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 324–329). ACM.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.
- Custard, M., & Sumner, T. (2005). Using machine learning to support quality judgments. *D-Lib Magazine, 11*(10). Retrieved from <http://www.dlib.org/dlib/october05/custard/10custard.html>
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105.
- Egerton, E. O., & Posey, L. (2007). *Quality standards inventory*. Retrieved from <http://learn.gwnursing.org/Education/QSI/QSI.html>
- Han, L., Neilands, T. B., & Dolcini, M. M. (2001). *Factor analysis of categorical data in SAS*. Retrieved from <https://www.lexjansen.com/wuss/2001/WUSS01044.pdf>
- Hixon, E., Barczyk, C., Ralston-Berg, P., & Buckenmeyer, J. (2016). The impact of previous online course experience students' perceptions of quality. *Online Learning, 20*(1), 25–40.

- Hoey, R. (2017). Examining the characteristics and content of instructor discussion interaction upon student outcomes in an online course. *Online Learning, 21*(4), 268–281.
- Illinois Center College. (2017). *Quality online course initiative*. Retrieved from <https://icc.edu/faculty-staff/files/ICC-QOCI-Version-4.4-published-05.04.17-2.pdf>
- Jaggars, S. S., & Xu, D. (2016). How do online course design features influence student performance? *Computers & Education, 95*, 270–284.
- Ke, F. (2013). Online interaction arrangements on quality of online interactions performed by diverse learners across disciplines. *Internet and Higher Education, 16*(1), 14–22.
- Kuo, Y., Walker, A. E., Belland, B. R., & Schroder, K. E. E. (2013). A predictive study of student satisfaction in online education programs in online courses. *International Review of Research in Open and Distance Learning, 14*(1), 1–39.
- Lee, J. (2014). An exploratory study of effective online learning: Assessing satisfaction levels of graduate students of mathematics education associated with human and design factors of an online course. *International Review of Research in Open and Distance Learning, 15*(1), 111–132.
- Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development, 59*(5), 593–618.
- Liu, I. F., Chen, M. C., Sun, Y. S., Wible, D., & Kuo, C. H. (2010). Extending the TAM model to explore the factors that affect intention to use an online learning community. *Computers & Education, 54*(2), 600–610.
- McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., and Bullock Mann, F. (2018). *The Condition of education 2018* (NCES 2018-144). U.S. Department of Education. National Center for Education Statistics.
- Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education, 3*, 1–7.
- Murray, M., Pérez, J., Geist, D., & Hedrick, A. (2012). Student interaction with online course content: Build it and they might come. *Journal of Information Technology Education: Research, 11*, 125–140.
- New Mexico State University (2011). *Online course design rubric*. Retrieved from <http://hpdr.education.nmsu.edu/files/2013/07/on-line-course-design-rubric-.pdf>
- Palomar College (2012). *Online course best practices checklist*. Retrieved from <http://www2.palomar.edu/poet/BestPracticesChecklistSP12.pdf>
- Quality Matters (2018). *Quality matters higher education rubric*. Retrieved from <https://www.qualitymatters.org/sites/default/files/PDFs/StandardsfromtheQMHigherEducationRubric.pdf>
- Raubenheimer, J. E. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology, 30*(4), 59–64.

- Reigeluth, C. (1999). *Instructional design theories and models: A new paradigm of instructional theory*. Lawrence Erlbaum Associates.
- Roblyer, M., & Wiencke, W. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *American Journal of Distance Education, 17*(2), 77–98.
- Sher, A. (2009). Assessing the relationship of student-instructor and student-student interaction to student learning and satisfaction in web-based online learning environment. *Journal of Interactive Online Learning, 8*(2), 102–120.
- Southern Regional Education Board. (2006). *Checklist for evaluating online courses*. Retrieved from <https://www.sreb.org/publication/checklist-evaluating-online-courses>
- Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*(4), 1183–1202.
- Swan, K., Matthews, D., Bogle, L., Boles, E., & Day, S. (2012). Linking online course design and implementation to learning outcomes: A design experiment. *The Internet and Higher Education, 15*, 81–88.
- University of North Dakota. (2016). *Rubric for evaluating online courses*. Retrieved from http://www1.und.edu/academics/center-for-instructional-and-learning-technologies/_files/docs/bestpracticesrubric.pdf
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79–94.
- Yuan, M., & Recker, M. (2015). Not all rubrics are equal: A review of rubrics for evaluating the quality of open educational resources. *International Review of Research in Open and Distributed Learning, 16*(5), 16–38.
- Yuan, M., & Recker, M. (2019). Does audience matter? Comparing teachers' and non-teachers' application and perception of quality rubrics for evaluating open educational resources. *Educational Technology Research and Development, 67*(1), 39–61.