Chen, L.-T., & Liu, L. (2018). Instructor's self-assessment of content design in online courses. International Journal of Technology in Teaching and learning, 14(1), 24-41.

Instructor's Self-Assessment of Content Design in Online Courses

Li-Ting Chen
Leping Liu
University of Nevada, Reno

There are only a handful of studies focused on assessing online teaching. Yet, the uniqueness of online course environment provides instructors more opportunities to conduct self-assessment of their own teaching than the traditional face-to-face classroom. In this paper, we demonstrate two types of self-assessment through two cases that online course instructors can use to evaluate the course content during their regular online instructional procedures. An end-of-semester oral examination was used in Case One whereas the weekly discussion posts were used in Case Two. The instructors performed self-assessment, based on which they redesigned the course contents and activities. Nonparametric analyses were used to examine students' learning outcomes. Results indicate that, in both cases, student learning improved with the redesigned contents and activities.

Keywords: Reflective teaching practices, self-study, distance learning, faculty, ADDIE model

INTRODUCTION

Assessing teaching effectiveness is important in higher education. Student evaluation (Berk, 2005; Felder & Brent, 2004; Kuzmanovic, Savic, Popovic, & Martic, 2013), peer evaluation or observation (Barnard et al. 2011; Berk, 2005; Drew & Klopper, 2014; Felder & Brent, 2004; Garcia, James, Bischof, & Baroffio, 2017; Hora & Ferrare, 2013), student learning outcome (Berk, 2005), student interview (Berk, 2005), videotaping teaching practice (Berk, 2005; Garcia et al., 2017), exit and alumni ratings (Berk, 2005), and reflection on teaching practice (Berk, 2005; Felder & Brent, 2004; Hora & Smolarek, 2018) are sources that can be used to inform teaching effectiveness as well as to evaluate areas that require improvement. While strategies for assessing teaching effectiveness in traditional face-to-face classrooms have been discussed extensively in the literature, there are only a handful of studies focused on assessing teaching effectiveness in online classrooms (Bangert, 2008; Ravenscroft, Luhanga, & King, 2017; Yang, 2017).

Li-Ting Chen is an Assistant Professor of Educational Measurements and Statistics in College of Education at University of Nevada, Reno. Leping Liu is a Professor of Information Technology and Statistics in College of Education at University of Nevada, Reno. Li-Ting Chen can be reached at litingc@unr.edu

Researchers have recommended different instruments based on student evaluation for assessing teaching effectiveness in traditional face-to-face classrooms and in online classrooms (Bangert, 2008; Ravenscroft, Luhanga, & King, 2017). In fact, the uniqueness of online teaching environment provides a variety of data sources for instructors to conduct self-assessment.

In this paper, we suggest instructors' self-assessment can be considered as reflective practice to self-evaluate the course content design in online classrooms. Results from the self-assessment may inform the instructors to implement change in the course design. Subsequent assessment of student learning can be used to evaluate whether the new course design is effective. Two cases are presented to share our experience of using assessment originally designed to evaluate student learning for self-assessing the course content design. Specifically, Case One presents the use of an end-of-semester oral examination during the final web or in-person conference for assessing student learning and self-assessing course content design. Case Two presents the use of weekly discussion posts as a self-assessment tool for (1) identifying student weakness, (2) linking of student weakness to course content, and (3) determining of modification needed. The new course design was expected to facilitate discussions that inspire deeper and critical thinking. Nonparametric analyses were used in the two cases to test whether the new course design facilitated student learning.

For the remainder of this paper, we review the literature on teaching effectiveness and evaluation tools for online teaching, principles for effective course design, and instructors' self-assessment. Details of the two cases are presented. The findings with respect to practice and directions for future research are discussed.

LITERATURE REVIEW

Online courses are more and more popular in higher education because online courses allow students to learn at their own pace and save money and time on traveling to campus. Based on the survey conducted by the National Center for Educational Statistics in 2007 (Parsad & Lewis, 2008), the top two factors that affected the decisions for offering online courses in higher education institutions are (a) meeting student demand for flexible schedules and (b) providing access to college for students who otherwise would not have access (e.g., because of geographic, family, or work-related reasons). Rao and Tanners (2011) reported students who took online courses favored the wide range of options and choices provided by online courses, such as the multiple formats of the learning materials. To address the popularity in online courses, it is important to understand what effective teaching is and what the evaluation tools are for assessing teaching effectiveness in online classrooms.

TEACHING EFFECTIVENESS AND EVALUATION TOOLS FOR ONLINE TEACHING

Seven principles of effective teaching recommended by Chickering and Gamson (1987) is one of the well-known summaries of research-based instructional practices. According to Chickering and Gamson (1987), effective teaching practices encourage: student-faculty contact, cooperation among students, active learning, prompt feedback, time on task, higher expectations, and respect for diverse talents and ways of learning.

Based on a qualitative study with focus group interviews of 11 nursing faculty members in an online program, Frazer, Sullivan, Weatherspoon, and Hussey (2017) identified five areas that define teaching effectiveness in online teaching. Specifically, the instructors in online learning classes need to (a) facilitate student learning, (b) aim to feel connected with students, (c) share experiences, (d) be approachable, (e) establish mutual comfort, and (f) be responsive to students' needs (Frazer et al., 2017, p. 3). In addition,

Frazer et al. (2017) suggested that timely feedback with a special emphasis on personal interaction is the key for teaching effectiveness. From graduate students' perspectives in online statistics courses, Yang (2017) concluded that effective online course design features include "clear course objectives, good alignment between course objectives and assessments, consistent module structure, a variety of assignments and learning activities, and a good balance between theory and applications" (p. 13). Yang (2017) found that case studies were rated by the students as the most effective instructional strategy among six instructional strategies adopted in these courses. The six instructional strategies were online discussion forum, video demonstrations of procedures in using a statistical analysis software, case studies of published research articles, mini projects, learning reflections, and other module assignments.

Because evaluation for online teaching include the technical and pedagogical aspects of the instruction, scholars have recommended different instruments to evaluate teaching effectiveness in online classrooms (Bangert, 2004, 2008; Sheard & Markham, 2005; Stewart, Hong, & Strudler, 2004; Ravenscroft, Luhanga, & King, 2017). For example, Bangert (2004, 2008) developed the Student Evaluation of Online Teaching Effectiveness (SEOTE) instrument based on Chickering and Gamson's (1987) seven principles of effective teaching. Through a series of validation studies, Bangert (2008) concluded that the 23 items of SEOTE can be used for assessing online teaching quality.

EFFECTIVE COURSE DESIGN

Effective teaching starts with intentional course design. The ADDIE model has been recommended for instructional design. ADDIE stands for the five phases of instructional design: **A**nalysis, **D**esign, **D**evelopment, **I**mplementation, and **E**valuation. The model was developed in 1970's and it is still commonly used for instructional design. The ADDIE model can be applied in order to close a gap between the acquired knowledge and the desired learning outcome.

In the analysis phase of the ADDIE model, an instructor determines whether or not instruction will close the gap and to what degree (Branch, 2009). Strategies to close the gap also needs to be recommended in the analysis phase (Branch, 2009). In the design phase, the instructor identifies the objectives and determines how the objectives will be met (Seels & Glasgow, 1998). The instructor also decides the instructional strategies that will be used to meet the objectives (Seels & Glasgow, 1998). It is crucial to determine how objectives will be assessed and what the assessment tools will be used in the design phase. The development phase is to "generate and validate the learning resources" (Branch, 2009, p. 3). The instructor in this phase can develop or use existed materials and media for their instruction. Formative evaluation is conducted in this phase to determine if students will learn from the learning resources and how the learning resources can be improved. In the implementation stage, the instructor needs to play an active role to "prepare for the learning environment and engage the students" (Branch, 2009, p. 3).

Evaluation is a multidimensional and essential component of the ADDIE model. Evaluation can be used in any phase of the ADDIE model, such as the formative evaluation in the development phase or the summative evaluation at the end of a course. In the evaluation phase, the instructor determines if the gap has been closed and if the objectives have been met. Furthermore, the instructor determines the impact of the learning resources and the changes required for the future delivery of the course (Peterson, 2003). Bentley and Kerhwald (2017) presented a case to support that sound pedagogical principles should be followed for successful teaching regardless of face-to-face or online teaching.

In this paper, evaluation of teaching effectiveness was reflected on course content design and was performed through the instructors' self-assessment. The literature has shown that self-assessment may be considered as a form of reflective practice.

INSTRUCTORS' SELF-ASSESSMENT

Dewey (1997) suggested reflective thinking involved two elements: "(a) a state of perplexity, hesitation, doubt; and (b) an act of search or investigation directed toward bringing to light further facts which serve to corroborate or to nullify the suggested belief" (p. 9). Schön (1983, 1987) defined two types of reflective practice: reflection-on-action and reflection-in-action. Reflection-on-action takes place when an instructor engages in an after-the-fact critical thought of practice of professionals. In contrast, reflection-in-action takes place when an instructor engages in a way of thinking about a situation while teaching, "in order to reframe and solve some breakdown in the smooth running of experience" (Ray-Bennett, Masys, Shiroshita, & Jackson, 2015, p. 104). Recently, Armstrong and Asselin (2017) wrote "reflective teaching practice is a process of self-examination and self-evaluation to gain insight into teaching to improve the teaching-learning experience" (p. 354). From a qualitative study of excellent instructors nominated by heads of departments in sciences, Kane, Sandretto, and Heath (2004) concluded that these instructors frequently engaged in purposeful reflective practice with respect to subject knowledge, pedagogical skill, interpersonal relationships, connections between research and teaching, and personality.

Instructors' self-assessment can be conducted before, during, and after delivering a course. Kennedy (2015) described using a Quality Matters rubric developed from technological pedagogical content knowledge (TPCK) to evaluate her own online teaching. Kennedy (2015) wrote "TPCK is *knowledge of technology to represent content and enable strategies*" (p. 149). Bentley and Kerhwald (2017) reported using reflective journals during the course development and while teaching the course to determine essentials for successful face-to-face teaching and online teaching. Garcia et al. (2017) evaluated an approach that problem-based learning tutors were videotaped for the tutoring sessions and the videos were later evaluated by the tutor themselves and peers. In addition, an instrument with 24 items regarding observable teaching behaviors or strategies was used to guide both the tutors' self-reflection processes and peer feedback. The results showed that tutors viewed self-observation and the rating on the 24 items providing opportunities to self-evaluate strategies they used to enhance student learning. Interestingly, 65% of the tutors who were interviewed one year later reported self-observation being more useful for them than receiving peer feedback and using the 24 item instrument.

From interviews of university instructors at three large, public universities in California, Hora and Smolarek (2018) found these instructors used both of numeric (i.e., final course grades, scores on assignments and exams, student course evaluations, peer evaluations) and non-numeric (i.e., pedagogical literature, teaching blogs, professional development materials, communication with colleagues and peers, direct communication with students, nonverbal cues of the students) data for reflective practice. Furthermore, these instructors' reflection can be categorized into instrumental reflection, structural-critical reflection, and social-critical reflection (Hora & Smolarek, 2018). Instrumental reflection is the reflection on day to day decision making of teaching activities, such as rearranging the course topic order. Structural-critical reflection is defined as reflection on contextual factors within the institutions, such as the ineffectiveness of the adopted learning management system. Social-critical reflection is referred to as "sociocultural issues that affected teaching and learning" (Hora & Smolarek, 2018, p. 566), such as poor performance of underrepresented student populations.

PURPOSES OF THE PAPER

The purpose of the paper is to demonstrate strategies for self-assessment through two cases that online course instructors may use to improve online teaching and student

learning outcome. Self-assessment is considered as reflective practice in the paper. Instead of the typical formal/numeric (e.g., student course evaluations) or informal/non-numeric (e.g. reflective journals) data for reflective practice, the two cases presented here used either an oral exam at the end of the semester or the weekly discussion posts throughout the semester to reflect on the course content design. Therefore, the reflection practice was reflection-on-action based on Schön (1983, 1987). The type of reflection was instrumental reflection according to Hora and Smolarek (2018).

CASE ONE: INSTRUCTOR'S SELF-ASSESSMENT BASED ON END-OF-SEMESTER ORAL EXAMS

This case presents our experiences to perform an instructor's self-assessment on the content design of online courses, through assessing students' learning based on their end-of-semester oral exams. Results of the self-assessment indicated areas that required improvement for the course content design and guided the new course design.

RESEARCH QUESTIONS

The case was guided by the following two research questions:

1. How does an instructor use the end-of-semester oral exam to conduct self-assessment on online teaching?
2. Do students better understand the conceptual knowledge related to educational research with modified learning materials and activities based on the instructor's self-assessment than those with the original course content?

RESEARCH PARTICIPANTS

The participants in this case consisted of students who took a fully online fundamental educational research methods course before (n_1 = 41) and after (n_2 = 31) implementing change in the course design. The course was offered in a western state university and designed to introduce the fundamental concepts, principles, and methods of educational research. Among the 41 students who took the course before revising the course design, 37 (90.2%) were in the master's degree program, two (4.9%) were in the doctoral degree program, and two (4.9%) were graduate special students who had not been admitted into graduate programs at the time they took the course. Among the 31 students who took the course after revising the course design, 26 (83.9%) were in the master's degree program, three (9.7%) were in the doctoral degree program, and two (6.5%) were graduate special students who had not been admitted into graduate programs at the time they took the course.

SETTINGS

The course used Canvas learning management system to deliver learning materials. The same instructor taught the course before and after the implementation of the new course design. Same assignments were assigned to all the students (N = 72). Each student was required to meet the instructor individually at least twice throughout the semester: in the middle and at the end of the semester. The first meeting was aimed to connect with the student and to collect early student feedback. In the second meeting, the instructor conducted an oral exam to assess student learning outcome. Similar to the first meeting, the instructor also collected student feedback in the second meeting. For both meetings, students were allowed to choose either to meet the instructor in person or online through BigBlueButton (an open-source web conference system).

PROCEDURES

As described above, originally, the first meeting was aimed to make connection with students and to receive early feedback whereas the second meeting was designed to assess student learning outcome and receive student feedback. Students were expected to show their understanding in principles for conducting human research, qualitative and quantitative research in education, and strategies to validate the research findings during the oral exams in the second (or final) meetings. The oral exams of the second meetings were videorecorded. During the second meetings, the instructor took notes about student responses to the questions and student feedback. The instructor also reviewed the videos for revising and modifying the course content design.

From students' responses (before change) to the questions asked during the oral exams in the second meetings, the instructor was aware that most students struggled with the concepts of reliability and validity. In addition, students also suggested audio lectures for facilitating their understanding. The instructor made decisions about changing the course design in two areas: (a) types of course learning materials, and (b) the purpose of the first meeting. The original course content design and the revised course design was described in details below:

Course Design before Implementing Change. Course learning materials were text-based. The readings were assigned for each week and the instructor's summaries of the readings were provided through PowerPoint files. Responses to weekly discussion questions were required every week to facilitate student understanding.

During the first meeting, students were asked if they have any questions about the learning materials. The instructor also invited student to provide feedback to the course. The format and logistics of the second meeting were discussed with the students.

Course Design after Implementing Change. In response to the recommendation of using audio lectures and students' difficulties in understanding reliability and validity, the instructor provided the links to three videos titled "How Results Can Be Misleading: Problems With Reliability and Validity" (Behar & Weierich, 2012), "Doing Qualitative Research" (Creswell, 2015) and "Increasing Validity in Qualitative Research" (Clark Pope, 2017) from SAGE Research Methods. The three videos ranged from 10 minutes to 35 minutes. The same text-based materials were provided. Similar to the course design before implementing change, students were required to respond to weekly discussion questions.

During the first meeting, the students were asked if they have any question about the learning materials. The students were also asked to explain the concepts of reliability and validity. If a student was not able to clearly define the two concepts, the instructor verbally explained the two concepts to the student. Regardless of students' responses to the question of reliability and validity, the instructor pointed out where to find the three video clips from Canvas and encouraged students to watch these videos. The instructor emphasized that it is important to understand the concepts of reliability and validity for this course.

Similar to what were done before implementing change, students were invited to provide their early feedback to the course in the first meeting. The format and logistics of the second meeting were also discussed with the students. Changes made to the course design were outlined in Table 1.

Table 1. *Changes in the Course Content Designs Before and After*

	Before	**After**
Type of learning materials	• Text-based learning materials.	• Text-based learning materials. • Instructional videos from SAGE research methods.

The purpose of the first meeting	• Answering students' questions. • Receiving early student feedback. • Discussing the logistics of the second meeting.	• Answering students' questions. • Receiving early student feedback. • Discussing the logistics of the second meeting. • Emphasizing the importance of reliability and validity in educational research and clarifying the two concepts. • Reminding students to watch the three instructional videos.

To prepare for the oral exam in the second meeting, all the students were asked to edit a PowerPoint file. The PowerPoint file needed to be submitted to the instructor at least one day before the scheduled second meeting time. During the second meeting, students may or may not be asked about the concepts of reliability and validity, depending on if the two concepts were clearly defined in the PowerPoint file. Students were allowed to look at their notes or textbooks when they were in the second meeting.

According to the textbook used for the course, "**reliability** means that scores from an instrument are stable and consistent" and "**validity** is the development of sound evidence to demonstrate that the test interpretation (of scores about the concept or construct that the test is assumed to measure) matches its proposed use" (Creswell & Gutterman, 2019 p. 158). The video "How Results Can Be Misleading: Problems With Reliability and Validity" (Behar & Weierich, 2012) explained inter-rater reliability, test-retest reliability, and internal consistency with respect to reliability. The same video also explained construct validity, external validity, and internal validity.

MEASUREMENTS

The two measurements used to identify improvement of students' understanding on conceptual knowledge were their understanding of (a) reliability and (b) validity. A student's understanding on reliability was coded as 1 if the student clearly explained reliability based on the definition provided by Creswell and Gutterman (2019) or at least one of the three definitions of reliability provided by Behar and Weierich (2012). A student's understanding on reliability was coded as 0 otherwise. The instructor coded the data for all the students.

A student's understanding on validity was coded as 1 if the student clearly explained validity based on the definition provided by Creswell and Gutterman (2019) or at least one of the three definitions of validity provided by Behar and Weierich (2012). A student's understanding on validity was coded as 0 otherwise. The instructor coded the data for all the students.

DATA ANALYSIS AND RESULTS

Research Question 1. The first research question is "How does an instructor use the end-of-semester oral exam as a strategy to conduct self-assessment on teaching effectiveness?" This case showed that the instructor used the assessment originally designed for evaluating student learning outcome to self-evaluate the course content design. That is, whatever the student couldn't provide answers to during the oral exam

showed the areas that needed to be strengthen. Because the oral exam was a type of synchronous assessment, in which the student and the instructor interacted with each other at the same time frame, the instructor may not recall all the student responses and feedback. This happened frequently when the instructor had other meetings right after the oral exam and when back-to-back oral exams were scheduled. To overcome this disadvantage of oral exams for self-assessment on teaching effectiveness, the instructor took notes during the oral exams and videotaped the oral exams.

In this case, the results of the oral exam helped the instructor to identify the gap between the acquired knowledge and the desired learning outcome (analysis phase of the ADDIE model). That was, most of the students couldn't verbally explain reliability and validity. The instructor identified two strategies to close the gap: (a) adding instructional videos to the learning materials, and (b) emphasizing the importance of understanding reliability and validity in the first meeting. The two strategies were implemented in the subsequent section of the same online course. Student learning outcomes before and after implementing change were compared to evaluate the effectiveness of the new course design (evaluation phase of the ADDIE model).

Research Question 2. The second research question is "Do students better understand the conceptual knowledge related to educational research with modified learning materials and activities based on the instructor's self-assessment than those with the original course content?" Conceptual knowledge was measured by whether or not the student can clearly define reliability and validity during the oral exam.

Two two-group independent-samples chi-square tests with a dichotomous response variable were performed to answer the research question. The two groups were students who took the course before and after implementing change in course content design. The first dichotomous response variable was whether or not the student could clearly defined reliability (1 = being able to define reliability in the oral exam, and 0 = not being able to define reliability in the oral exam). The second dichotomous response variable was whether or not the student could clearly defined validity (1 = being able to define validity in the oral exam, and 0 = not being able to define validity in the oral exam).

The null hypothesis stated that students who took the course before and after implementing change in the course content design had the same distribution on the two dichotomous variables, namely being able to define reliability and validity. Table 2 presents the observed frequencies for the number of student who were able and not able to define reliability, given they took the course before or after the implementation of new course design. The two-group independent-samples chi-square test rejected the null hypothesis of equal distribution on being able to explain reliability for students who took the course before and after implementing change ($\chi^2 = 10.23$, $df = 1$, $p = .001$). Odds ratio was reported as a measure of effect size. The odds ratio of 4.95 (=(30*20)/(11*11))showed that students who took the course after the change were about five times more likely to define reliability correctly than students who took the course before the change.

Table 2. *The 2 by 2 Contingency Table on Reliability*

		Being able to explain reliability		
		No	Yes	Total
Group	Before change	30 (73.2%)	11 (26.8%)	41
	After change	11 (35.5%)	20 (64.5%)	31

Table 3 presents the observed frequencies for the number of student who were able and not able to define validity, given they took the course before or after the implementation of the new course design. The two-group independent-samples chi-square test failed to reject the null hypothesis of equal distribution on being able to explain validity for students who

took the course before and after implementing change (χ^2 = 2.38, *df* =1, *p* = .123). The odds ratio of 2.10 (=(26*17)/(14*15)) showed that students who took the course after the change were about two times more likely to define validity correctly than students who took the course before the change.

Table 3. *The 2 by 2 Contingency Table on Validity*

		Being able to explain validity		
		No	Yes	Total
Group	Before change	26 (63.4%)	15 (36.6%)	41
	After change	14 (45.2%)	17 (54.8%)	31

Overall, the odds ratios showed that students who took the course after the change were more likely to define the two concepts correctly than students who took the course before the change. Yet, only the null hypothesis that students who took the course before and after implementing change in course content design had the same distribution on understanding reliability was rejected.

CASE TWO: INSTRUCTOR'S SELF-ASSESSMENT BASED ON STUDENTS' ONLINE DISCUSSIONS

CASE CONTEXT AND RESEARCH QUESTIONS

This case presents our experiences to perform an instructor's self-assessment on the content design of online courses, through assessing students' learning based on their reflective online discussions. Specifically, from students' online discussions on certain assigned topics and tasks, the instructor locates where in the content design needs to improve, and simultaneously generates the way to modify.

Two online courses offered to education-major graduate students in a western state university are included in this case. One course, referred as Course A, focuses on the *design of online teaching and learning*, and the other one referred as Course B focuses on the *methods of using information technology in education*. Both courses are small size hands-on design courses and require students to learn and apply the fundamental knowledge and models of instructional design. They are offered only in spring semesters.

Our experiences of self-assessment were conducted with these two courses in two successive years. As shown in Table 4, the two courses taught in the first year (A1 and B1) served as Group 1 (n_1 = 8 + 6 = 14), the two courses taught in the second year (A2 and B2) served as Group 2 (n_2 = 7 + 5 = 12), and a total of 26 students served as the participants of this case.

Table 4. *Case Participants*

	Course A	**Course B**	Total
Year 1	8 (A1)	6 (B1)	14 (Group 1)
Year 2	7 (A2)	5 (B2)	12 (Group 2)
Total	15	11	26

Next, we present the case with the logistic procedures of the instructor's self-assessment. Research questions guided through this case were:

1. How does an instructor use students' online discussions to perform self-assessment on the content design of an online course?

2. Do Year 2 students better understand instructional design with modified learning materials and activities based on the instructor's self-assessment than Year 1 students with the original course content?

ASSESSING STUDENT LEARNING THROUGH ONLINE DISCUSSIONS

The Coursework. Both courses required students to understand and apply the principles and models of instructional design to create a technology-based *product* for teaching and learning. In Course A, each student was required to create an online course with at least two fully implemented lessons on a selected subject area (as individual project), deliver the course through an online learning system–Canvas, then "teach" the lessons online (as a group project), and at the end evaluate each other's design and teaching.

The product from Course B was an interactive segment of web-based or computer-based application for learning, such as (a) interactive drills, exercises, or instructions, (b) a game segment, (c) a hyperlinked interactive concept map, or (d) a chapter of a multimedia or hypermedia e-book. Each student was required to complete a product, integrate it into a real world teaching-learning environment, test and evaluate his/her work, and revise the product to put it into another new cycle of applications.

The coursework in both courses covered three common areas. First, instructional design theories and applications were the fundamental knowledge for the design of either the online course or the computer-based applications. The ADDIE model (Branch, 2009) was used as the framework to develop the products. Theme focused weekly discussions, projects and research on the five phases of the ADDIE model were required. Second, technology skills necessary to develop those products were introduced and practiced with weekly lab activities. The third area was the integration of design and use of technology into the content materials for a particular lesson or computer-based application. The final products was implemented through a comprehensive procedure upon each student's uniqueness of design.

Online Discussions. In both courses, there were ten theme-focused online discussion assignments. The themes included needs assessment, learner assessment, objective-oriented design, content analysis, task analysis, storyboarding, beta testing of the products, peer evaluation, lessons learned, and personalized design models.

For each theme, the instructor assigned reading materials (chapters, articles, or any online resources), listed the concepts, theories, models, procedures, or examples related to the objectives set for the learning on each theme, provided the discussion guidance with open-ended questions. Students were required to read the materials, write a reflective journal addressing the questions, then do a search to find at least two more articles or resources on the theme or at a different but relevant direction from the theme, and raise two questions for others to think. To complete the assignment, students posted their journals including the new resources and questions, read others posts, answered or discussed on the new questions others raised.

Assessing Student Learning from Online Discussion. The purpose of the online discussion assignments was for students to demonstrate whether or to what extent they understood the theme basics, planned appropriate tasks or procedures to achieve the goal of their designs, or improved any weak area during the rolling discussions in the class. In the course content design, the themes were all connected toward the quality of each of the A-D-D-I-E projects, and the quality of the final product. Therefore, the instructor was able to assess the progresses of learning through the on-going discussions, and meanwhile the instructor could also conduct self-assessment on the course content design through the online discussions.

INSTRUCTOR'S SELF-ASSESSMENT AND CONTENT REDESIGN

Common Procedures. The instructor's self-assessment can be conducted through three procedures. It starts with identifying the weakness in student learning, then continues to analyze where the weakness linked to in the course content, and ends with the redesign or modification of the related course content. From the online discussions on a given theme, the instructor can find some common weakness of students' learning that could be improved if certain additional materials, instructions, or activities were provided.

Table 5 summarizes the initial effort made to conduct an self-assessment: (a) identified a list of common weakness found in the online discussions of the two courses in the first year, (b) located where they pointed to in the contents, and (c) determined what kind of redesign or modification could be done. Our experiences of self-assessment are introduced next.

Table 5. *Instructor's Self-Assessment on the Content Design*

Student Weakness (*Identify*)	**Where in Contents** (*Link to*)	**Redesign/Modification** (*Implement*)
• Missing components • Incorrectly address the question • Not in right direction • Definition is not clear • Too open to reach a conclusion • Lack of design support • Logic is not reasonable • Example does not fit • Slow in progress	• Theme/topic • Procedures • Exercises • Activities • Models • Resources	• Add current literature • Modify reading materials • Provide real-world cases • Provide simple/clear examples • Provide visual structure • Promote group activities • Encourage more practice • Multi-level communications • Add more procedural contents

Case Procedures. We used two themes of the online discussions as examples for this case, as they were required in both courses. In the first year, the instructor identified some weakness from the online discussions in Group 1 courses. Then the instructor traced back to the course contents where the instructions on related themes were given. Lined up with the weaknesses, the instructions on all related themes were redesigned, and additional materials, examples, and guidance were provided. For example, the weaknesses on the themes of *Learner Assessment* and *Task Analysis,* the content links to the weaknesses, and the relevant modifications on these two themes are shown in Table 6.

Table 6. *Two Examples of Self-Assessment*

Student Weakness (*Identify*)	**Where in Contents** (*Link to*)	**Redesign/Modification** (*Implement*)
Learner Assessment		
• Not clear what to assess • Missing explanations on the assessment procedures • Not clear on the specific needs of the target audience • Not well summarized	• The phase of *Analysis* in the ADDIE model • Reading materials • Exercise procedures	• Provide current literature • Provide resources of instrument • Provide testing examples • Provide writing examples

Task Analysis		
• Lack of connections to the objectives • No detailed procedures of completing certain tasks • No detailed descriptions on materials, equipment or tools needed to complete certain tasks • Lack of resources for the task completion	• The phase of *Design* in the ADDIE model • Components of task analysis • Guidance to write the task analysis	• Provide visual structure of the design • Provide guidance to analyze individual tasks • Provide a set of operational verbs used to write the task analysis • Provide writing samples

In the second year, the modified version of course contents were offered to the Group 2 courses. The instructor did the assessment on student learning as usual and the self-assessment again on content design. The learning outcomes were compared to determine whether students improved their learning with the modified course contents.

MEASUREMENTS AND DATA COLLECTION

In this case, we compared the learning outcomes between Group 1 and Group 2 courses on the two theme discussions (*Learner Assessment*, and *Task Analysis*), and one *Project of Design*. The scores assigned to the students by the instructor were used to measure student learning.

The measurement on the theme discussion of *Learner Assessment* was a 10-point scale, to examine five required components: (a) definition of the learners, (b) some main characteristics or needs of the learners, (c) learner assessment procedures, (d) learning style or cognitive development features with at least one testing results using a self-developed instrument, or published instrument, and (e) a clear summary that can be used to guide the design of the final product (the online course with two completed lessons, or the computer-based interactive applications for learning).

The measurement on the theme discussion of *Task Analysis* was a 15-point scale, to examine five required components: (a) list of the tasks, (b) connection between each task and the objectives of product, (c) necessary materials, tools, resources or anything related, (d) criteria of quality for each task, and (e) timeline.

The discussions on these two themes were closely related to the quality of one of the project of the course – the project for the *Design* phase of ADDIE model. So we also included the measurement on this project. A 30-point scale was used to examine the components mainly on: (a) content design, (b) the design of technology use, and (c) the design of integration.

Scores obtained from the three measurements for all students were collected from the first year Group 1 courses and the second year Group 2 courses for further data analysis.

DATA ANALYSIS AND RESULTS

Research Question 1. The first research question is "How does an instructor use students' online discussions to perform self-assessment on the content design of an online course?"

This case presented here demonstrated that instructors can use the assessment methods originally designed to evaluate student learning outcome to conduct self-assessment on the course content design. An initial effort can start with the three procedures to conduct instructor's self-assessment with students' on-going online discussions: (a) assessing

student learning, (b) instructor's self-assessment, and (c) redesign or modifications on the content.

The first is to assess student learning through theme focused online discussions. We set clear learning objectives and developed measurements for each theme. During the online discussion, if some weaknesses (as listed in Table 6) appeared twice or more, we marked them for further self-assessment on the content design. If some weaknesses appeared but solved during discussions, we marked them for further self-assessment on the activity design. If some weaknesses appeared because of students' lack of interest, motivation, or preparations, we marked them for further self-assessment on the design of objective-oriented production. From this procedure, we obtained first hand qualitative data for our self-assessment.

In the second procedure – the instructor's self-assessment, each weakness appeared in the online discussions were lined out and mapped back to the content design. We found that the mapping did make it clear where or how we could improve our content design. Therefore, followed in the third procedure – redesign and modification, we had clear targets and criteria for redesign, were aware of the exact contents or activities to be modified, and were able to set more accurate standards or requirements for student learning (see Table 5 for some general examples).

Research Question 2. The second research question is "Do Year 2 students better understand instructional design with modified learning materials and activities based on the instructor's self-assessment than Year 1 students with the original course content?" To answer this question, we compared the quantitative measures on *Learner Assessment*, *Task Analysis*, and the *Project of Design* between students in Group 1 courses (with original content materials) and Group 2 courses (with modified content materials).

As we used a small sample size of 26 with unequal size groups, and the tested scores could not assume a normal distribution, we chose a nonparametric statistical method Mann-Whitney *U* for the data analysis (Cohen, 2001; Conover, 1999). The number of times that a randomly selected score from one sample is higher than the randomly selected score from the other compared sample is called the Mann-Whitney *U* statistics (Grissom & Kim, 2012, p. 151). Three Mann-Whitney *U* tests were conducted to test the difference in the ranks of the learning outcome scores between students who learned with original content materials (in Group 1 courses) and those who learned with modified content materials (in Group 2 courses). The null hypothesis for the Mann-Whitney *U* tests was that there is no tendency of difference in the ranks of learning outcome scores between students who learned with original content materials and those who learned with modified content materials. Alpha level at .05 was set for the tests.

In the first Mann-Whitney *U* test, the scores of student learning on the theme *Learner Assessment* from Group 1 courses and Group 2 courses were compared. Using the ranked scores, the results indicated a significant difference between the two groups ($U = 32$, $n_1 = 14$, $n_2 = 12$, $p = .006$). The sum of ranks for Group 2 courses ($\Sigma R_2 = 214$) was higher than the sum of ranks for Group 1 courses ($\Sigma R_1 = 137$). Moreover, the effect size (ES) for the group difference was 0.529, calculated with the formula $ES = |Z| / \sqrt{N}$, as suggested by Corder and Foreman (2014, p. 80).

In the second Mann-Whitney *U* test, the scores of student learning on the theme *Task Analysis* from Group 1 courses and Group 2 courses were compared. Using the ranked scores, the results indicated a significant difference between the two groups ($U = 43$, $n_1 = 14$, $n_2 = 12$, $p = .036$). The sum of ranks for Group 2 courses ($\Sigma R_2 = 203$) was higher than the sum of ranks for Group 1 courses ($\Sigma R_1 = 148$). The ES for the group difference was 0.416.

In the third Mann-Whitney *U* test, the scores of student learning on the project of *Design* from Group 1 courses and Group 2 courses were compared. Using the ranked

scores, the results indicated a significant difference between the two groups (U = 44, n_1 =14, n_2 =12, p = 0.041). The sum of ranks for Group 2 courses (ΣR_2 = 202) was higher than the sum of ranks for Group 1 courses (ΣR_1 = 148). The ES for the group difference was 0.406.

Overall, in all three tests, the null hypothesis can be rejected that there is no tendency of difference in the ranks of learning outcome scores between students who learned with original content materials and those who learned with modified content materials. Therefore our results support that students' learning can be improved with the modified content materials, and the instructor's self-assessment based on the learning outcomes from students' online discussions can contribute a positive input to the online course content design.

DISCUSSIONS

This paper presents two different types of instructors' self-assessment in online courses: (a) Case one: self-assessment based on using the end-of-semester oral exam, and (b) Case two: self-assessment using the weekly discussion posts. In Case One, the data collected for self-assessment were similar to the use of exam results for reflective practice (Hora & Smolarek, 2018). In Case two, online discussion posts were used for self-assessment. Although we did not find literature on using online discussion posts for self-assessment, this was similar to use student assignments for reflective practice (Hora & Smolarek, 2018). The two cases provide evidence that the online learning environment offers unique opportunities for instructors to conduct self-assessment of their teaching. More importantly, the self-assessment may lead to necessary implementation of new course design.

The entire process of the instructors' self-assessment corresponds to the five phases of the ADDIE model–**A**nalysis, **D**esign, **D**evelopment, **I**mplementation, and **E**valuation (Brancg, 2009). That is, through the self-assessment, the instructor acknowledges the learning gap and formulates strategies to close the learning gap–analysis. The instructor then designs and develops workable teaching materials and strategies that meet the objectives. Lastly, the instructor implements the new course design and evaluates student learning through nonparametric statistics. For the areas that do not show significant improvement, the instructor may implement a new course design and restart another ADDIE cycle.

Based on what we learned from the two cases, we would like to share the following conclusions and thoughts with our readers.

DATA-DRIVEN DECISION MAKING

Issues of accountability, assessment, and data-driven decision making have gained a growing attention in the field of higher education (Cox et al., 2017; Hora & Smolarek, 2018; Lane, 2014). It is suggested that administrators and faculty should make decisions based on assessment data. The decisions are expected to enhance student learning experience. The two cases presented in this paper are applications of data-driven decision making for improving student learning based on the instructors' reflective practice.

With today's technology, researchers also recommended big data analytics in higher education for data-driven decision making. Big data analytics examines large amounts of data that enables researchers to identify patterns and correlations. Student data collect by the Office of Institutional Office may provide opportunities to understand hidden patterns and complex correlations, and predict student success. For example, Lane (2014) wrote "student ID cards allow institutions to track library usage, dining habits on and off campus,

health center usage, medical treatments, attendance at cocurricular activities, and what residents halls or classrooms students access" (xviii-xix). Indeed, the use of learning management system, such as Canvas, in online courses enables instructors to collect meaningful data from individual students. These data include but not limit to time to login the system, time spent on reading the learning materials, and interaction with other students. Future studies may continue exploring the use of learning management system for instructors' self-assessment and decision making in online courses.

INSTRUCTORS' SELF-ASSESSMENT OF ASSESSMENT

The two cases presented in this paper were originally designed for assessing student learning outcome. We then examined and explored using the results of student assessment as the instructor's self-assessment. We called this as the assessment of assessment or meta-assessment (Liu & Chen, 2018). In contrast to the data collected from students (e.g., student course evaluation) or peers (e.g., peer observation), the instructor can use meta-assessment to get the first-hand data. The first-hand data immediately inform the instructor about areas that require improvement.

Once again, the unique online learning environment provides multiple opportunities for self-assessment as reflective practice. The current study only presents two examples of self-assessment in online courses. Other data may be used for an instructor's self-assessment. Researchers can explore using other types of data for self-assessment of online teaching. For example, frequent communication with students to clarify the requirement of an assignment may indicate a need to revise the guidance of the assignment.

MECHANISM OF IMPROVEMENT ON STUDENT LEARNING OUTCOME

After implementing the new course design, it showed learning improvement for students. What was the mechanism to enhance student learning using the new course design? We suggest Vygostsky's zone of proximal development (ZPD) can be used for explaining the improved student learning. Vygostsky (1978) wrote ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86).

We suggest that the revision and the modification of the course design may enhance the learning experience through interaction with peers, the instructor, or the artifacts. Specifically, in Case One, the instructor asked students to explain the concepts of reliability and validity in the first meeting after implementing change. When students couldn't clearly explain the two concepts, the instructor illustrated the two concepts to the students. The instructional videos on reliability and validity were also provided to facilitate student understating of reliability and validity. In Case Two, students' responses to the discussion question was influenced by the other students' replies and the new course learning materials. The instructor also replied to the discussion question to maintain constructive discussion. The interaction with peers, the instructor, or the artifacts (new course learning materials) therefore increased the developmental level on course related knowledge.

LIMITATION AND FURTHER STUDIES

There were two limitations of the study. First, because the data were collected and analyzed by the instructors who themselves taught the courses, one may argued that the results were biased. Future research may replicate our cases with multiple coders for the data. Second, only two indicators were used to measure conceptual knowledge in Case

One, and these two indicators had their own limitation. In addition to conceptual knowledge, future studies may consider factual knowledge, procedural knowledge, and metacognitive knowledge (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) for self-assessment.

The two cases presented in the paper used nonparametric statistics to evaluate student learning before and after implementing the new course design. Compared to parametric statistics that can only be used for interval or ratio data, nonparametric statistics can be used when data are collected from nominal or ordinal scale. Nonparametric can also be used when assumptions such as normal distribution and equal variances are not met. Lastly, the graduate level courses are usually small class sizes with $N < 30$. When studies are conducted using graduate level courses, nonparametric statistics may be required even when the data are collected using interval scale instrument. As Pett (2016) wrote "sample size requirements also are less stringent for nonparametric tests. It is not unusual for sample sizes of 20 or less to be reported" (p. 4). Future research may employ nonparametric statistics to analyze similar data when the use parametric statistics is not appropriate.

We hope that the two cases presented in this paper can promote high quality of online teaching and facilitate more discussions on using self-assessment as a strategy for improving online teaching.

REFERENCES

Armstrong, D. K., & Asselin, M. E. (2017). Supporting faculty during pedagogical change through reflective teaching practice: An innovative approach. *Nursing Education Perspectives, 38* (6), 354–357.

Bangert, A. W. (2004). The seven principles of good practice: A framework for evaluating online teaching. *The Internet and Higher Education, 7*(3), 217–232.

Bangert, A. W. (2008). The development and validation of the student evaluation of online teaching effectiveness. *Computers in the Schools, 25*, 25–47.

Barnard, A., Croft, W., Irons, R., Cuffe, N., Bandara, W., & Rowntree, P. (2011). Peer partnership to enhance scholarship of teaching: A case study. *Higher Education Research & Development, 30*, 435–448.

Behar, E., & Weierich, M. (2012). *How results can be misleading: problems with reliability and validity* [Streaming video]. Retrieved from SAGE Research Methods.

Bentley, B., & Kerhwald, B. A. (2017). From 'good teaching' to 'better teaching': One academic's journey to online teaching. *Journal of Perspective in Applied Academic Practice, 5*(1), 58–66.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education, 17*, 48–62.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York, NY: David McKay.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York, NY: Springer.

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J. P., & King, F. J. (1975). Interservice procedures for instructional systems development. Executive summary and model. Florida State University, Tallahassee Center for Educational Technology.

Chickering, A. W., & Gamson, Z. F. (1987, March). Seven principles for good practice in undergraduate education. *AAHE bulletin, 39*(7), 3–7.

Clark Pope, D. (2017). *Increasing validity in qualitative research*[Streaming video]. Retrieved from SAGE Research Methods.

Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York: John Wiley and Sons.

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: John Wiley and Sons.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach* (2nd ed.). Hoboken, NJ: Wiley.

Cox, B. E., Reason, R. D., Tobolowsky, B. F., Brower, R. L., Patterson, S., Luczyk, S., & Roberts, K. (2017) Lip service or actionable insights? Linking student experiences to institutional assessment and data-driven decision making in higher education, *The Journal of Higher Education, 88*(6), 835–862.

Creswell, J. (Academic). (2015). *Doing qualitative research* [Streaming video]. Retrieved from SAGE Research Methods.

Creswell, J. H., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Upper Saddle River, NJ: Pearson Education.

Dewey, J. (1997). *How we think* (An unabridged republication). Mineola, NY: Dover Publications, Inc.

Drew, S., & Klopper, C. (2014). Evaluating faculty pedagogic practice to inform strategic academic professional development: A case of cases. *Higher Education, 67*, 349–367.

Felder, R. M., & Brent, R. (2004). How to evaluate teaching. *Chemical Engineering Education, 38*(3), 200–202.

Frazer, C., Sullivan, D. H., Weatherspoon, D., & Hussey, L. (2017). Faculty perceptions of online teaching effectiveness and indicators of quality. *Nursing Research and Practice, 2017*, 1–6.

Garcia, I., James, R. W., Bischof, P., & Baroffio, A. (2017). Self-observation and peer feedback as a faculty development approach for problem-based learning tutors: A program evaluation. *Teaching and Learning in Medicine, 29*(3), 313–325.

Grissom, R. J., & Kim, J. J. (2012). *Effect size for research: Univariate and multivariate applications* (2nd Ed), New York: Routledge: Taylor & Francis Group.

Hora, M. T., & Ferrare, J. J. (2013). Instructional systems of practice: A multidimensional analysis of math and science undergraduate course planning and classroom teaching. *The Journal of the Learning Sciences, 22*, 212–257.

Hora, M. T., & Smolarek, B. B. (2018). Examining faculty reflective practice: A call for critical awareness and institutional support. *The Journal of Higher Education, 89*, 553–581.

Kane, R., Sandretto, S., & Heath, C. (2004). An investigation into excellent tertiary teaching: Emphasizing reflective practices. *Higher Education, 47*(3), 283–310.

Kennedy, J. (2015). Using TPCK as a scaffold to self-assess the novice online teaching experience. *Distance Education, 36*(1), 148–154.

Kuzmanovic, M., Savic, G., Popovic, M., & Martic, M. (2013). A new approach to evaluation of university teaching considering heterogeneity of students' preferences. *Higher Education, 66*, 153–171.

Lane, J. E. (Ed.). (2014). *Building a smarter university: Big data, innovation, and analytics*. Albany, NY: SUNY Press.

Liu, L. & Chen, L.-T. (2018). Conducting synchronous assessment through web videoconference to improve online learning: Case outcome with nonparametric analysis. *Journal of Educational Technology Development and Exchange, 11*(1), 45–64.

Parsad, B., & Lewis, L. (2008). *Distance education at degree-granting postsecondary institutions: 2006–07* (NCES 2009–044). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Peterson, C. (2003). Bringing ADDIE to life: Instructional design at its best. *Journal of Educational Multimedia and Hypermedia, 12*(3), 227–241.

Pett, M. A. (2016). *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Rao, K., & Tanners, A. (2011). Curb cuts in cyberspace: Universal instructional design for online courses. *Journal of Postsecondary Education and Disability, 24*(3), 211–229.

Ravenscroft, B., Luhanga, U., & King, B. (2017). Adapting Bangert's online teaching effectiveness evaluation tool to a Canadian context. *Innovation in Education and Teaching International, 54*, 355–363.

Ray-Bennett, N. S., Masys, A., Shiroshita, H., & Jackson, P. (2015). Reactive to proactive to reflective disaster responses: Introducing critical reflective practices in disaster risk reduction. In A. E., Collins, S. Jones, B. Manyena, & J. Jayawickrama. (Eds.). *Hazards, risks, and disasters in society* (pp. 99–117).

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.

Seels, B., & Glasgow, Z. (1998). *Making instructional design decisions* (2nd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Sheard, J., & Markham, S. (2005).Web-based learning environments: Developing a framework for evaluation. *Assessment and Evaluation in Higher Education, 30*(4), 353–368.

Stewart, I., Hong, E., & Strudler, N. (2004). Development and validation of an instrument for student evaluation of the quality of web-based instruction. *The American Journal of Distance Education, 18*(3), 131–150.

Vygotsky, L. S. (1978). *Mind and society*. Cambridge, MA: Harvard University Press.

Yang, D. (2017). Instructional strategies and course design for teaching statistics online: perspective from online students. *International Journal of STEM Education, 4*(34), 1–15.