

# Journal of Educational Technology Development and Exchange (JETDE)

---

Volume 11 | Issue 1

---

12-15-2018

## Conducting Synchronous Assessment through Web Videoconference to Improve Online Learning: Case Outcomes with Nonparametric Analysis

Leping Liu

University of Nevada, Reno, liu@unr.edu

Li-Ting Chen

University of Nevada, Reno, litingc@unr.edu

Follow this and additional works at: <https://aquila.usm.edu/jetde>



Part of the [Instructional Media Design Commons](#), [Online and Distance Education Commons](#), and the [Other Education Commons](#)

---

### Recommended Citation

Liu, Leping and Chen, Li-Ting (2018) "Conducting Synchronous Assessment through Web Videoconference to Improve Online Learning: Case Outcomes with Nonparametric Analysis," *Journal of Educational Technology Development and Exchange (JETDE)*: Vol. 11 : Iss. 1 , Article 4.

DOI: 10.18785/jetde.1101.04

Available at: <https://aquila.usm.edu/jetde/vol11/iss1/4>

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Journal of Educational Technology Development and Exchange (JETDE) by an authorized editor of The Aquila Digital Community. For more information, please contact [Joshua.Cromwell@usm.edu](mailto:Joshua.Cromwell@usm.edu).

## Conducting Synchronous Assessment through Web Videoconference to Improve Online Learning: Case Outcomes with Nonparametric Analysis

---

**Leping Liu**

University of Nevada, Reno

**Li-Ting Chen**

University of Nevada, Reno

**Abstract:** *Online assessment has always been a challenge to online teaching. Educators have been exploring a variety of methods to perform online assessment. However, it appears that there is not enough work in the field focusing on online synchronous assessment. This paper presents two cases that demonstrate the design and implementation of using web videoconference for synchronous assessment in an educational research methods online course and an instructional video production online course. The purpose of the two cases was to explore whether or with what methods student online learning could be improved through synchronous assessment. Case outcomes were analyzed with nonparametric methods, and the results did show students' improvement in their learning, specifically in their understanding and mastering of factual, conceptual, procedural and metacognitive knowledge. Methods, procedures, tips and cautions of conducting such videoconference-based synchronous assessment in online courses are discussed.*

**Keywords:** Synchronous Assessment, Online Learning, Web Videoconference, Knowledge Taxonomy, Nonparametric Analysis

### 1. Introduction

Online courses are offered in almost every university and college in the United States (Holzweiss, Joyner, Fuller, Henderson, & Young, 2014; Liu, Ripley, & Lee, 2016; Scott, Temple, & Marshall, 2015), and approximate 27% of students of public higher education institutions take at least one distance course (Allen, Seaman, Poulin, & Straut, 2016). Over

years, instructors have devoted a tremendous amount of work on online assessment (Barber, King, & Buchanan, 2015), exploring a variety of ways such as online quizzes or online homework (Lowe, 2015), online discussions (Kent, Laslo, & Rafaeli, 2016; Klisc, McGill, & Hobbs, 2009), online peer evaluation (Alvarez, Espasa, & Guasch, 2012), online videoconference (Bower, 2011; Dymont & Downing, 2018a; Okada, & Scott,

2015), or using learning analytics to assess student online learning (Martin & Ndoeye, 2016; Nyland, Davies, Chapman, & Allen, 2017). However, designing and conducting efficient online assessment has always been a challenging area to instructors (Atherton, Shah, Vazquez, Griffiths, Jackson, & Burgess, 2017; Cheng, Jordan, Schallert, & D-Team, 2013; Dennen, 2008).

Basically, online assessment can be sorted into two main types: synchronous assessment and asynchronous assessment, pending on whether a real-time online interaction between

the instructor and learner(s) is conducted. Most methods and activities presented in the literature are in the scope of asynchronous assessment; it appears that there is not enough research and practice in the field on online synchronous assessment (Chao, & Hung, Chen, 2012; Lee & Liu, 2016). Figure 1 shows the trends of research interests in online assessment (Figure 1A) and learning from assessment (Figure 1B) from 2013 to present. The same as in the literature, in a search by “synchronous assessment” or “online synchronous assessment”, there is not enough data to show the trend in the field.

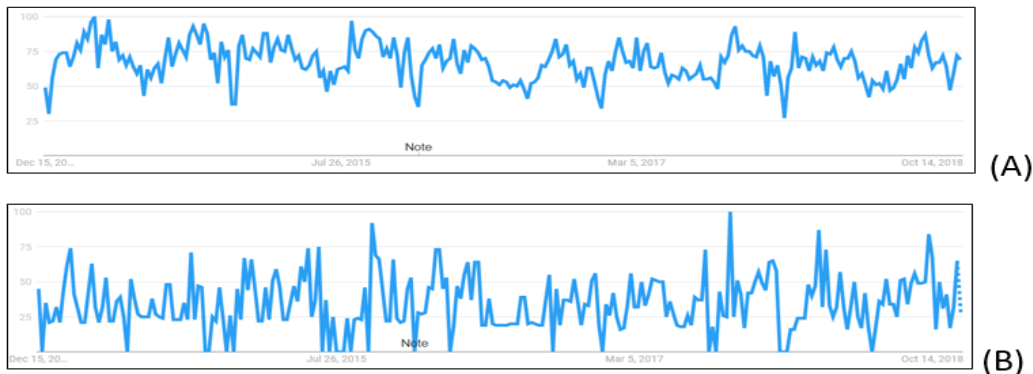


Figure 1. Trends of online assessment (A) and learning from assessment (B)

(Google Trends, 2018)

This paper presents two cases that demonstrate the design and implementation of using web videoconference to conduct synchronous assessment in an educational research methods online course and an instructional video production online course. The purpose of the cases is to explore whether or with what methods student online learning could be improved through synchronous assessment, and specifically to what extent their understanding and mastering of factual, conceptual, procedural and metacognitive knowledge could be improved. Case outcomes are analyzed with nonparametric methods.

## 2. Literature Review

### 2.1. Design of Online Assessment

A successful online course is built upon tremendous amount of work on course design and delivery (Conole, Dyke, Oliver, & Seale, 2004; Jaggars & Xu, 2016). Design of assessment starts at the stage of course design, following the ADDIE instructional design model that lines out the main principles in the five phases of instructional design: *Analysis, Design, Development, Implementation, and Evaluation* (Branson, Rayner, Cox, Furman,

& King, 1975; Gagne, Wager, Golas, & Keller, 2005). Design of assessment mainly includes the decisions on the following components:

*Learning Objectives and Outcomes.* This is a major outcome from the phase of *Analysis*. It provides a foundation or framework for activities in the phase of *Design*, where all decisions on content information, pedagogy strategies, learning activities and assessment plan can be made (Gagne et al., 2005; Liu & Velasquezbryant, 2003).

*Learning Contents.* Content analysis is performed according to the learning objectives and outcomes (from the phase of *Analysis*), and the range, level, order, format, and delivery methods of content information will be determined in the phase of *Design*. Then, content materials such as lecture notes, learning materials, reading references, and video lessons or tutorials will be completed during the phase of *Development* (Dini & Liu, 2017; Liu & Gibson, 2018). Also, technology preparations for online teaching and learning is another huge task in the phase of *Development*.

*Learning Activities.* Along with the above activities, corresponding online learning activities will be planned with operational procedures and to-do list. Tasks, procedures, outcomes and evaluation criteria for individual work and collaborative group work will be clearly described. This is part of the work done in the phase of *Development*. (Dini &

Liu, 2017; Mundkur & Ellickson, 2012).

*Assessment Plan.* The purpose of assessment is to determine whether or to what extent students' performances and learning outcomes meet the expected criterion set in the objectives. Pending on the types of activities or learning outcomes, we can choose the type of assessment: synchronous or asynchronous assessment (Liu, 2018; Liu & Gibson, 2018; Liu & Johnson, 2002). Online synchronous assessment is conducted with real-time online interactions between the instructor and students or among students; asynchronous assessment is performed with activities that are not simultaneous or concurrent in time (Chao et al., 2012; Lee & Liu, 2016). With any type, an assessment plan includes methods, measurements and instrument, activities to conduct the assessment, technology tools, and timing (Liu, 2018; Liu & Johnson, 2002). The cases in this paper demonstrate some practical methods the authors used to conduct online synchronous assessment.

## 2.2. The Knowledge Taxonomy – FCPM

To better measure and assess learning, we may specify the measures and assessment methods on different dimensions of knowledge. Theoretically, knowledge is taxonomically classified into four dimensions or categories: **f**actual, **c**onceptual, **p**rocedural and **m**etacognitive (FCPM), by Bloom, Engelhart, Furst, Hill, and Krathwohl (1956) and Anderson et al. (2001).

Table 1. Summary of FCPM Knowledge Taxonomy

Dimensions of FCPM Knowledge Taxonomy (Anderson et al., 2001)	Assessment Methods
<b>Factual Knowledge:</b> the basic units of knowledge that the learners must know in a discipline. e.g., terminology and basic concepts.	Multiple choice quizzes (Mayotte, 2010)
<b>Conceptual Knowledge:</b> interrelations among the basic units within structure. e.g., classifications and categories, principles and generalizations, theories, models, and structures.	Essay-based testing (Foltz, Laham, & Landauer, 1999)

Table 1. Continued Summary of FCPM Knowledge Taxonomy

Dimensions of FCPM Knowledge Taxonomy (Anderson et al., 2001)	Assessment Methods
<b>Procedural Knowledge:</b> explaining how to perform a task, and goal-oriented methods of inquiry, criteria for using skills, algorithms, techniques, and methods.	Demonstration, hands-on projects (Garris, Ahlers, & Driskell 2002)
<b>Metacognitive Knowledge:</b> awareness and knowledge of one’s own cognition – knowledge about knowledge. It involves strategies for learning, thinking, and problem solving.	Oral presentations, applying discursive oral, dialogue, or other communication methods (Cohen, 1998; Nakatani, 2005)

Table 1 summarizes the definitions and examples of the FCPM knowledge taxonomy, and assessment methods proposed in the literature for each (Chao et al., 2012). Factual and conceptual knowledge are about “knowing what” and they are considered as the basic units of knowledge. Procedural and metacognitive knowledge are about “knowing how” and they are referred to as more comprehensive and higher level of knowledge (Chao et al., 2012; Mayer, 2002). The next question is how the different dimension of knowledge taught in an online course could be assessed with appropriate online synchronous assessment methods.

**2.3. Online Synchronous Assessment, Potential Challenges and Possible Solutions**

All assessment methods listed in Table 1 could be performed as synchronous assessments and asynchronous assessments. Synchronous assessment can be done in the format of text and voice chat, telephone conversations, videoconferencing, or even meetings in virtual spaces such as Second Life, where discussions can be facilitated among groups of students (Nakatani, 2005; Uribe & Vaughan, 2017). Chao, Hung, and Chen (2012) reported their study on four online synchronous assessment methods: (a) using synchronous quizzes to assess factual knowledge, (b) using synchronous practices

to assess procedural knowledge, (c) using synchronous essays to assess conceptual knowledge, and (d) and using synchronous oral exams to assess metacognitive knowledge. Those are very practical examples of using particular methods to assess each dimension of the FCPM knowledge taxonomy. As described in the 2.1 section above, online synchronous assessments have completed the assessment procedures, decisions, and tasks in four ADDIE phases of *Analysis, Design, Development, and Implementation* (Gagne et al., 2005).

In educators’ experiences, two main potential challenges are recognized when conducting such synchronous assessment. The first challenge is the lack of appropriate technology platforms or tools that enable or support the synchronous assessment from a full-dimension real-time interaction between learners and the instructor (Liu & Gibson, 2018). For example, in a traditional videoconference or a cyber-classroom, the learner’s video screen can only provide the evidence that he/she is really the learner who is supposed to participate the online synchronous examinations. The instructor would not see any type of “helper” or information resources that the learner might have out of the focus and range of the Web camera.

One possible solution could be the

Meeting OWL, an intelligent 360° all-in-one video conferencing device. It has three features: (a) dynamic video and audio: the 360° lens automatically focuses on the users as they speak, while 8 omnidirectional microphones locate the presenter or speakers with clear audio; (b) all in one: it combines video and audio to simplify the meeting setup, like a dynamic and sophisticated webcam; and (c) no installation: it connects via USB, operates software free, and is compatible with any web-based meeting platform (OWL, 2018). This device would enable the instructor to “see” all the dimensions where a learner is assessed.

The second challenge when conducting an online synchronous assessment is that different types of subjects and courses need different assessment methods that require different functional and technology supports for the assessment activities (Hayes & Ringwood, 2008; Liu & Gibson, 2018; Zlatović, Balaban, & Kermek, 2015). While there is no simple solution for this, the decisions on content design, content preparation, design of learning activities, and assessment design would straighten out the operation list and required technology support.

In the following sections, we present two cases that demonstrate the design and procedures to perform online synchronous assessments on student learning of:

1. factual and conceptual knowledge in an online educational research methods course (Case One), and
2. procedural and metacognitive knowledge of design in an online course on instructional video production (Case Two).

Nonparametric analyses are used to assess the outcomes of the online synchronous assessment.

### **3. Case One: Synchronous Assessment on Factual and Conceptual Knowledge of Educational Research Methods**

#### **3.1. Research Questions**

In this case, we explored the use of synchronous assessment (oral exam) to assess student learning outcome. The case was guided by the following two research questions:

1. Can the synchronous assessment method (oral exam) be used to better assess student learning outcome than the asynchronous assessment method (traditional written exam)?
2. What do we learn from using such an oral exam to assess student learning outcome?

#### **3.2. Participants**

Participants ( $N = 41$ ) consisted of students from two sections of a fully online fundamental educational research methods course offered in a western state university in the United States. The course was designed to introduce basic statistics concepts, and general process and practice of educational research for graduate students. Among the 41 students, 37 were in the master’s degree program, two were in the doctoral degree program, and two graduate special students who had not been admitted into graduate programs at the time they took the course.

#### **3.3. Settings**

The course was taught through an online learning management system *Canvas*. The same instructor taught both online sections. The course requirements included (a) responding to weekly discussion questions, (b) completing six homework assignments,

(c) acquiring the Collaborative Institutional Training Initiative (CITI) training certificate, (d) completing one quantitative article analysis paper and one qualitative article analysis paper, and (e) attending two conferences to meet with the instructor individually.

The first conferences were held in the middle of the semester, aiming to connect with students, verify the identity of students, and receive early feedback from students. The second conferences were held at the end of semester, aiming to assess student learning outcome, and receive student feedback for further improvement of the course design and teaching. Canvas Scheduler tool was used to create time slots, and all students signed up for their individual conferences with the instructor. Students were allowed to choose whether they wanted to meet the instructor in person or online through an online conference program *BigBlueButton*.

### **3.4. Procedures**

Online assessment for this case took place in the forms of both asynchronous and synchronous assessment:

*Asynchronous Assessment.* Students received a study guide two weeks before the final conferences and were asked to create a PowerPoint file to prepare for the final conferences. The PowerPoint file was required to present the definition of educational research, and similarities and differences between quantitative and qualitative research. Students were asked to summarize the similarities that focus on (a) the reasons why research is important, (b) the process of research, (c) basic principles for conducting human research, and (d) principles of a good research report. In addition, students were asked to address the differences based on (a) the research questions, (b) the types of data, (c) ways to analyze the data, (d) strategies

to validate the findings, and (e) researcher-researched relationship. This PowerPoint file was the written exam as a form of asynchronous assessment. Students submitted it to the instructor the day before the final conferences.

*Synchronous Assessment.* The synchronous assessment was conducted through oral exams during the final web conferences (or in person conferences as preferred). At the beginning of the final conferences, permission to video record the oral exam was obtained from all students. During the final conferences, the instructor asked students questions for additional clarification based on the content of the student's PowerPoint file. For example, a student listed *reliability and validity* under quantitative research without any further explanation. The instructor would ask the student to explain and redefine *reliability and validity*. If the required component was missing on the PowerPoint file, the instructor would prompt the student to address it. For instance, if the student did not list any strategy for validating findings in qualitative research, the instructor would ask if the student could think of any ways that the subjects/interviewees may help a qualitative researcher to validate his/her research findings (i.e., member checking). It was the instructor's hope that this synchronous assessment could improve student learning.

### **3.5. Measurements**

Eight indicators (see Table 2) regarding reliability and validity in educational research were coded for both asynchronous assessment (the traditional written exam using PowerPoint files), and synchronous assessment (the oral exam during final conferences). An indicator was coded as 0 if the student did not mention or clearly explain the concept, or 1 if he/she

clearly explained the concept.

Table 2. Eight Indicators for Reliability and Validity (concepts)

Indicators	Description
Reliability	“ <b>Reliability</b> means that scores from an instrument are stable and consistent” (Creswell & Guetterman, 2019, p. 158)
Validity	“ <b>Validity</b> 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 & Guetterman, 2019, p. 158)
Internal/External Validity	Internal validity: the extent to which a cause-and-effect inference can be correctly drawn. External validity: the extent to which the study results can be generalized to the target population.
Confirmability	“Researchers can address confirmability (the qualitative counterpart to bias) by admitting biases and assumptions and acknowledging limitations in the study’s methods” (Creswell & Guetterman, 2019, p. 261)
Triangulation	“Qualitative inquiries triangulate among different data sources to enhance the accuracy of a study” (Creswell & Guetterman, 2019, p. 261)
Member checking	“A process in which the researcher asks one or more participants in the study to check the accuracy of the account” (Creswell & Guetterman, 2019, p. 261)
External audit	“Researchers may also ask a person outside the project to conduct a thorough review of the study and report back, in writing, the strengths and weakness of the project” (Creswell & Guetterman, 2019, p. 262)
Transferability	“Transferability (external validity) from one setting to another can be established by establishing the context of a study, giving detailed descriptions of the procedures and writing findings in vivid detail supported with quotes” (Creswell & Guetterman, 2019, p. 261)

The eight indicators are in the scope of *factual and conceptual* knowledge as defined in Table 1. They can be measured with both written and oral exams. Scores on the written exam indicated the initial measures on each indicator for each student. Then the eight indicators were coded again as the final measures based on students’ responses to the instructors’ questions during the final conferences. Again, when the

student responded to the instructor’s prompt with clearly explanation for the concept, the corresponding indicator was coded as 1. When the student didn’t provide correct explanation for the concept, the corresponding indicator remained 0. For *each indicator*, the asynchronous-synchronous (written-oral exam) scores could be in one of the following combinations:



- 0 – 0: when the coding for a specific indicator based on the written exam (the PowerPoint file) was 0, the coding based on the student’s responses to the instructor’s questions during the final conference (the oral exam) could be 0 if the student still did not provide the expected answers or responses.

- 0 – 1: when the coding for a specific indicator based on the PowerPoint file was 0, the coding based on the student’s responses to the instructor’s questions during the final conference could be 1 if the student was able to provide clear explanations for the concept.

- 1 – 1: when the coding for a specific indicator based on PowerPoint file was 1, the coding based on students’ responses during the final conference remained 1.

Values for the eight indicators were added together to indicate the number of concepts regarding to reliability and validity in educational research recalled by the students when they were assessed based on

the asynchronous assessment (the initial written exams using PowerPoint files) and when they were assessed through synchronous assessment (the final oral exams).

Table 3 presents the frequency distribution for the number of students who recalled from zero to six indicators during the final conferences, grouping by the initial number of indicators clearly presented in the PowerPoint file. No student could recall more than six indicators either on the PowerPoint file or during the final conferences. According to Table 3, for example, among the 24 students who did not clearly explain or identify any of the eight indicators on the PowerPoint file (with a score of 0), six students (or 25.0%) were still not able to recall any of the reliability and validity concepts during the final conferences, six students (25.0%) recalled one more indicator, eight students (33.3%) recalled two more indicators, three students (12.5%) recalled three more indicators, and one student (4.2%) recalled four more indicators.

*Table 3. The Number of Students (percentages) Who Recalled Zero to Four Additional Indicators During Final Conferences, Grouping by the Initial Number of Indicators Presented in the PowerPoint Files*

No. indicators presented on the ppt file	No. additional indicators explained during final conferences					Total
	0	1	2	3	4	
0	6 (25.0)	6 (25.0)	8 (33.3)	3 (12.5)	1 (4.2)	24 (100.0)
1	1 (12.5)	4 (50.0)	1 (12.5)	2 (25.0)	0 (0)	8 (100.0)
2	2 (33.3)	3 (50.0)	1 (16.7)	0 (0)	0 (0)	6 (100.0)
3	2 (100.0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (100.0)
6	1 (100.0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (100.0)

*Note.* No student clearly presented four, five, seven, or eight indicators on the PowerPoint file. No student clearly explained five or more additional indicators during the final conferences. Based on our coding strategy, it was not possible that one student recalled less number of the indicators during the final conference than on the PowerPoint file.

### 3.6. Data Analysis and Results

*Research Question 1.* The first research question is “can the synchronous assessment method (oral exam) be used to better examine student learning outcome than the asynchronous assessment method (traditional written exam)?” The PowerPoint files prepared by the students for the final conferences were treated as the outcomes from traditional written exams whereas the student responses to the instructor’s questions during the final conferences were treated as the outcomes from the oral exams. The total number of indicators clearly explained in the PowerPoint file and the total number of indicators clearly explained during the final conference were compared for each student. Each student was then coded as

0 – for not being able to explain and define *additional* indicators during the final conference,

1 – for being able to explain and define *at least one additional* indicators during the final conference.

On the basis of pure chance, equal number of students should be observed for not being able to explain and define additional indicators during final conferences and for being able to explain and define at least one additional indicators during final conferences. The logic is that when the results showed that the number of students who were able to explain and define at least one additional indicators during final conferences was much

greater than the expected number due to chance, it implied that oral exams can be used to facilitate students’ thinking process, and explore in depth how much a student does understand and has learned. In such case, it supported the hypothesis that the oral exam can be used to better examine student learning outcome.

A nonparametric method, called the one-way chi-square test, was carried out to test the null hypothesis that only chance determined the number of students who could or could not clearly explained additional number of indicators during the final conference. Alpha level at .05 was used for the test.

Table 4 presents the observed and expected frequencies for the number of students who were not able to explain and defined additional indicators during final conferences and for the number of students who were able to explain and defined at least one additional indicators during final conferences. The one-way chi-square test rejected the null hypothesis of only chance determined the number of students in these two groups ( $\chi^2 = 7.05$ ,  $df = 1$ ,  $p = .008$ ). Twelve out of the 41 students (29%) were not able to recall additional indicators during the final conferences, and 29 students (71%) were able to recall at least one additional indicator during the final conferences.

*Research Question 2.* The second research question is “what do we learned from using oral exams to assess student learning outcome?” From this case experience, we want to share several tips that may be of

Table 4. The Observed and Expected Frequencies Table (N = 41)

Did not recalled additional indicators	Recalled additional indicators
Observed = 12 (29%) Expected = 20.5 (50%)	Observed = 29 (71%) Expected = 20.5 (50%)

reference to other educators who are interested in conducting the similar synchronous assessment.

First, test audio and video and have a backup plan. The instructor had students who were more willing to meet her in person than online and these students explained that either they liked face time with the instructor or they were not comfortable with technology. Therefore, we suggest whenever the in-person meeting is possible, instructors should offer the option of in-person meeting to the students. When web videoconferences were conducted, the instructor did experience trouble hearing or viewing the student. We suggest that instructors should remind students to test the audio and video settings before conducting the web videoconferences. To prevent cancellation of the web videoconferences from possible technical issues, instructors may provide the phone number for students to call in.

Second, inform the instructor for rescheduling. The Canvas Scheduler was used for signing up for the two conferences. A few students rescheduled their conferences on the same date or one day prior his/her original scheduled time. To prevent instructors from missing the conferences, it is critical to have the students who reschedule the conferences also send a message to the instructors. Instructors should also turn on notifications for appointment cancellations and for student appointment signups.

Finally, reserve a quiet room for the conferences. During the web videoconferences, the instructor heard noises such as dog barking and child crying. To ensure the quality of web videoconferences, instructors may request students to choose a quiet room where the web videoconference can be conducted without interruption.

#### **4. Case Two: Synchronous Assessment on Procedural and Metacognitive Knowledge in Instructional Video Design**

##### **4.1. Case Context and Research Question**

This case presents our experiences of using web conference and a content analysis program MAXQDA to assess the quality of instructional video productions that students created in an online course. The online video production course was offered to education students in a western state university. Two sections of the online course with a total of 24 students were included in this case.

We presented Case One in the format of a traditional research report. Case Two is presented along with the operational procedures to conduct this particular synchronous assessment. Research questions guided through this case were:

1. Can the synchronous assessment method (oral exam) be used to improve the quality of students' instructional video design?
2. To what extent can students' procedural and metacognitive knowledge in instructional design be evaluated with the synchronous assessment method?

##### **4.2. What to Assess?**

The purpose of the course was to prepare students with the knowledge and skills to develop image-based instructional materials, or video productions. In this course, students learned

1. the history, major issues and trends to use image-based materials or video products to improve teaching and learning;
2. technology integration theories and instructional design models;

3. skills and technology tools used to create such visual products; and

4. the methods to apply the video products created in this course in a real world learning context.

The visual productions students created could be graphics, images, pictures, and video clips focusing on a learning topic. All the pieces of products were then used to create an instructional video (see Figure 1). The instructional videos included video lessons



Figure 1. Sample instructional videos

created by teacher education graduates on the topics of history, biology, geography, science, algebra, calculus, and other subjects. The length of the videos varied from 15 minutes to 40 minutes.

The major components to be assessed in this case was the knowledge and skills of instructional design, focusing on the procedural and metacognitive knowledge (as defined in Table 1); the procedures of using the design principles in video production, and students' self-awareness of their use of those design principles.

#### 4.3. How to Assess?

Evaluating the quality of instructional videos starts from setting the evaluation criteria. According to the principles of design in ADDIE model (Analysis, Design, Development, Implementation and Evaluation) by Gagne et al. (2005), the twelve criteria were set to evaluate the quality of the instructional videos:

1. statement of goals/objectives,
2. description of the audience,
3. topic introduction and prerequisite knowledge of skills,

4. structure of how the lesson is organized,
5. content information,
6. logic of content flow,
7. examples,
8. assessment,
9. appropriate media uses,
10. frame transitions,
11. screen captions, and
12. other criteria pending on the subject areas.

*Asynchronous Assessment* was conducted by the instructor over time, evaluating the quality of students' video development at each stage of the production. Students posted their in-processing product on the course discussion board every week. Weekly feedbacks from the instructor and classmates were provided to students for the continual improvement of their production.

*Synchronous Assessment* was performed twice during the semester for two purposes: (a) to assess the procedural knowledge (the quality of students' video development), and

(b) to assess the metacognitive knowledge of their design (the extent to which they are aware of the instructional design principles they used in their work. Two individual web videoconferences were conducted).

The first conference was at the midpoint of their video production, to assess the developmental progresses and quality of the video production. One score was recorded on the quality of the video at the time. A qualitative review of the progresses in the video production was discussed with students.

The second conference was at the final of the semester. The quality of the final video product was assessed with a final score. A qualitative review and evaluation was discussed with students on their awareness of the design principles they used.

#### 4.4. Synchronous Assessment with Web Conference and MAXQDA

*BigBlueButton* was chosen to conduct web conferences in this online course. It is the same online conference tool with the functions to conduct one-to-one, one-to-many, and

many-to-many online videoconferences, text-based chatting, and online presentations to share files and resources.

A content analysis program MAXQDA was used to evaluate and visualize the quality of students' video products. It is a proprietary, professional software package for qualitative and mixed methods data management and analysis (MAXQDA, 2014). The software can perform content analysis on a text file and on a video file.

*Procedures of Coding and Analyzing.* To analyze the video quality, first, each of the design quality criterion was coded into a color in MAXQDA. Colors can be used to present certain pattern, for example, if we are coding skills, then a darker color code could indicate a more advanced skills. Alternatively, if we are coding qualities, then a group of similar colors (e.g., light blue, to dark blue) could indicate quality criteria under a certain stage (e.g., analysis stage, development stage, or implementation stage of the ADDIE model). Figure 2 is the color code for the 12 design quality criteria, and the video segments being marked with the criteria.

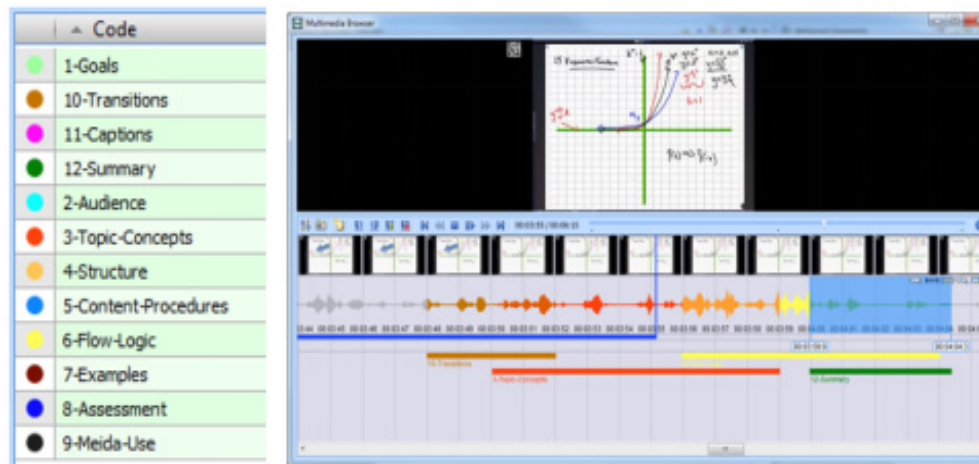


Figure 2. Coding of the video quality criteria to be evaluated

When analyzing the video, we imported the movie/video file into MAXQDA, then started to play it. During the play, we first examined what design criterion a segment of movie demonstrated (e.g., the statement of the purposes, or it demonstrates the logic of the theory, or knowledge), then dragged the coding criterion to tracks under the movie (the right part of Figure 2), and extended the color code to the same length as the length of frames crossing the movie with that criterion. Sometimes, a segment of movie may show multiple criteria of the quality design, then multiple color codes can be paralleled or overlapped.

*Data Visualization.* MAXQDA then generated the data visualization graphics (Figure 3). Looking at the color distribution of each 30X40 matrices, the quality of a video product, the quality distribution throughout the video can be easily described or interpreted. The frames, or the video moments that met certain of the 12 criteria were summarized in a color matrices. For example, the left part of Figure 3 showed that the ongoing project was lack of design quality. It met certain of the criteria, but needed to improve in those empty places. The right part of the Figure 3 showed a complete video lesson with improved high design qualities, as all the video moments demonstrated certain quality of the design.

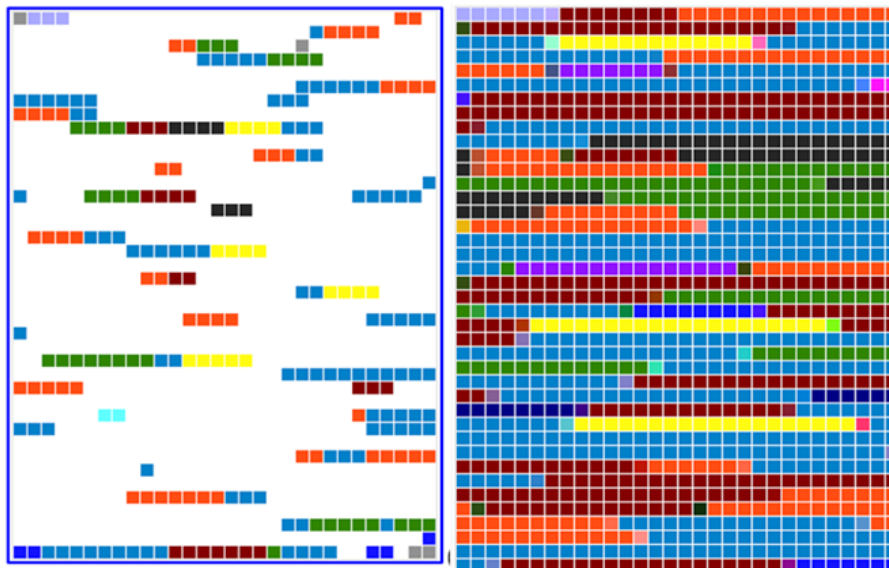


Figure 3. Visualizing video quality

*Conducting Synchronous Assessment with Web Conference.* Individual web conferences were then scheduled and conducted. Over the conference, the quality of the student's ongoing visual product, and the color matrices that shows what quality is missing at which part of the lesson, and what might be done to improve were shared and discussed with the

students. Figure 4 demonstrates the screen sharing area at the web conference.

#### 4.5. Measurements and Data Collection

Both quantitative data and qualitative data were collected through the two web conferences as shown in Table 5.

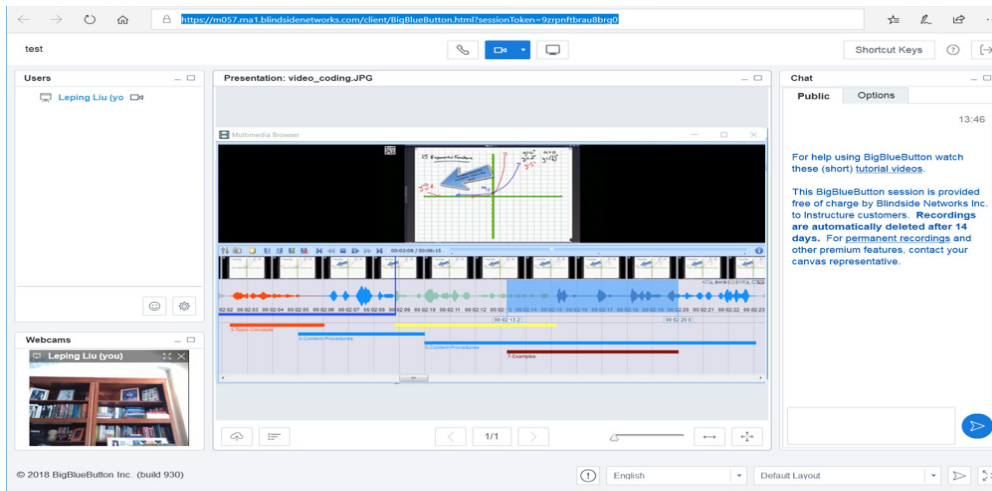


Figure 4. Video analysis shown in the screen sharing area of the web conference

*Quantitative Data.* The 12 criteria were used to measure the design quality of the instructional video products. Two quantitative scores were obtained as evaluation one and evaluation two scores. Evaluation one score was given at the midpoint of the video

production while first web conference was conducted, and evaluation two score was given at final while the second web conference was conducted. Scores ranged from 0 to 12. One point was given when one criterion was well demonstrated in the video.

Table 5. Summary of Data Collection

	Quantitative Data	Qualitative Data
Conference 1	Evaluation 1: Mid-video quality	Procedural knowledge review
Conference 2	Evaluation 2: Final-video quality	Metacognitive knowledge review

*Qualitative Data.* The qualitative review of the progresses in the video production (the procedural knowledge) was discussed at the first conference, focusing on (a) if the video production was progressed as planned, (b) if the quality was achieved as expected, and (c) if the timeline was reasonably followed.

video that met the criteria, (b) describe what design principles from the ADDIE model were demonstrated in certain segments of their video, and (c) explain what might be done differently if creating a similar video in the future.

The second qualitative review, on students' self-awareness of the design principles they used (the metacognitive knowledge), was conducted at the final conference. Students were asked to (a) review the qualities of their

In these two qualitative reviews, basically, the instructor asked questions and discussed with the students on each focus, and the conversations were recorded for further documentation.

#### 4.6. Data Analysis and Results

*Research Question 1* examines whether the synchronous assessment method (oral exam) can be used to improve the quality of students' instructional video design. The logic is that (a) synchronous assessment was used in a real time interaction with students and obtained the two scores of video quality, (b) if students' scores in evaluation two was significantly increased comparing with the scores in evaluation one, we may consider that the evaluation and suggestions from the first conference did provide students more directions to improve their video product, and hence (c) we may view it as one of the positive outcomes from the synchronous assessment procedures, through which student learning was improved.

As we used a small sample size of 24, and the tested scores could not assume a normal distribution, we chose a nonparametric statistical method for the data analysis (Cohen, 2001; Conover, 1999). Simply, a sign test was conducted to test the difference of the medians between the two related data (scores from evaluation one and evaluation two). The null hypothesis for the sign test was that the median scores between evaluation one and evaluation two should be equal, due to pure chance. Alpha level at .05 was set for the test.

Results presented that the sign test  $R$  was significant ( $R = 4$ , and  $p = .004$ ), so the null hypothesis was rejected, and the median differences between the two evaluation scores were significant. Among the 24 students, 18 (75%) increased their scores in evaluation two, 2 (8%) tied in the two evaluations, and 4 (17%) received less scores in evaluation two. Overall, the results showed that the median score of evaluation two was significantly higher than that of evaluation one, indicating the improvement of student learning.

*Research Question 2* aims to explore the extent to which students' procedural knowledge and metacognitive knowledge in instructional design can be evaluated with the synchronous assessment method. First, in evaluating students' procedural knowledge in instructional design we found that (a) student understanding or mastering of the knowledge in instructional design were progressed with the procedures of their video production, (b) the quality of the video evaluated at the time reflected whether their decisions were appropriate in the initial design, (c) adjustment of the decisions could be done during the video production, and (d) the timeline is another critical factor to ensure the completion of the products.

Second, in evaluating students' metacognitive knowledge in instructional design we found that (a) the design quality is consistent with students' self-awareness of the instructional design principles they used, for example, if they could clearly explain the instructional design principles applied in the design, the related quality scores would be higher, and those who received less scores in evaluation two did have difficulties to clearly express their self-awareness of the design principles; (b) students' self-awareness of the principles enabled them to perform a re-evaluation of their work after the completion of their work, which is another improvement of learning; and (c) students' self-awareness of the principles also enabled them to formulate some different approaches of design that may better their design in future work and made them realize a variety of options in design generated from the original ADDIE model, such as standard-based design by subject areas, design of collaborative team production, or design of the interaction with the users of their instructional video product.



## 5. Discussions and Conclusions

In summary, the two cases demonstrated: (a) different approaches to conduct synchronous assessment in an online course, (b) the method of using real-time web videoconference to obtain both quantitative and qualitative assessment data, (c) the method to assess each category of the taxonomically classified knowledge: factual and conceptual knowledge in Case One and procedural and metacognitive knowledge in Case Two, and (d) the method of utilizing nonparametric statistical methods to analyze data collected from such un-randomized small sample. From what we learned from our case experiences, we would like to share the following conclusions and thoughts with our colleagues and readers.

*Improve Learning from Synchronous Assessment.* The original purpose of assessment is to find out whether or to what extent student learning has achieved the course objectives with expected quality. The results and findings from the two cases also suggest another “function” of assessment: to improve learning through assessment. Well-designed online synchronous assessment can provide the opportunities for instructors to (a) assess student learning, and (b) deliver additional instructions or guidance in a real-time one-to-one conversation and help students improve their learning. In the literature, consistent findings are found from Dymont and Downing’s (2018b) study, which has the similar positive outcomes, using weekly web conference to facilitate and improve student online learning.

*Start Assessment from Design.* Planning of assessment starts from the original course design, for example, the original content design and the design of technology use. First, as in these two cases, the procedures,

activities and methods of synchronous assessment were determined along with the results from the content analysis at each category of knowledge offered in the online courses: factual knowledge (Mayotte, 2010), conceptual knowledge (Foltz et al., 1999), procedural knowledge (Garris et al., 2002), and metacognitive knowledge (Cohen, 1998; Nakatani, 2005). Second, the selection of technology tools or platforms for assessment need to carefully integrate the content design into the design of technology use (Liu & Velasquezbryant, 2003). In our cases, web videoconference was the tool that could be used to better assess student learning.

*Consider Nonparametric Statistical Methods.* The two cases also demonstrated the methods of using nonparametric tests to analyze data from such a small sample while the normality assumption is not met (Cohen, 2001; Conover, 1999). Very often, such data could not be meaningfully used to produce solid guidance to current or further practice. For instance, in some manuscripts we read before, such data were either inappropriately analyzed with parametric tests, or simply treated with descriptive analysis (Liu, 2015; Liu, Gibson, & Maddux, 2013). We hope the data analysis methods introduced in this paper could provide an example as a possible solution to this situation. For example, the results from the two cases clearly exhibited the differences: (a) with the chi-square test, we found that the proportion of students who recalled more research design indicators was significantly higher than the proportion of those who did not recall more indicators, and (b) with the sign test we found that the median score of video production in evaluation two was significantly higher than that in evaluation one. Such findings at least could provide a reason for further explorations with an experimental design and larger size of sample.

*Limitations and Further Studies.* The two cases mainly focus on our practical experiences. There are some limitations that can be addressed in future studies. First, the literature in certain theories such as the theory of knowledge taxonomy was not explored in depth. If the features of the taxonomically classified knowledge were explored and described in more details, readers may have more understanding on the content-related assessment decisions. Second, the online synchronous assessment on the procedural and metacognitive knowledge of instructional design were conducted with a qualitative review; only qualitative data were collected. We may continue our work and conduct studies with quantitative measures on procedural and metacognitive knowledge. Further studies also can be developed to assess the effectiveness of the online synchronous assessment, that is, to perform meta-assessment.

We hope our experiences in these two cases can be of reference to other educators who have the similar interest in online synchronous assessments. Comments and suggestions are appreciated.

## References

- Allen, E, Seaman, J., Poulin, R., & Straut, T. T. (2016). Online report card: Tracking online education in the United States. Retrieved from <https://eric.ed.gov/?id=ED572777>
- Alvarez, I., Espasa, A., & Guasch, T. (2012). The value of feedback in improving collaborative writing assignments in an online learning environment. *Studies in Higher Education*, 37 (4), 387-400. doi:10.1080/03075079.2010.510182
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Atherton, M., Shah, M., Vazquez, J., Griffiths, Z., Jackson, B., & Burgess, C. (2017). Using learning analytics to assess student engagement and academic outcomes in open access enabling programs. *Open Learning: The Journal of Open, Distance and e-Learning*, 32(2), 119-136. doi:10.1080/02680513.2017.1309646
- Barber, W., King, S., & Buchanan, S. (2015). Problem based learning and authentic assessment in digital pedagogy: Embracing the role of collaborative communities. *The Electronic Journal of e-Learning*, 13(2), 59-67.
- 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.
- Bower, M. (2011). Redesigning a web-conferencing environment to scaffold computing students' creative design processes. *Journal of Educational*

- Technology & Society*, 14(1), 27-42. Retrieved from <https://www.j-ets.net/ETS/index-2.html>
- 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.
- Chao, K. J., Hung, I. C., & Chen, N. S. (2012). On the design of online synchronous assessments in a synchronous cyber classroom. *Journal of Computer Assisted Learning*, 28, 379-395. doi:10.1111/j.1365-2729.2011.00463.x
- Cheng, A.-C., Jordan, M. E., Schallert, D. L., & D-Team, T. (2013). Reconsidering assessment in online/hybrid courses: Knowing versus learning. *Computers & Education*, 68, 51-59. doi:10.1016/j.compedu.2013.04.022
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. New York, NY: Addison-Wesley Longman.
- Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York: John Wiley and Sonse.
- Conole, G., Dyke, M., Oliver, M., & Seale, J. (2004). Mapping pedagogy and tools for effective learning design. *Computers & Education*, 43, 17-33.
- Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: John Wiley and Sons.
- 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.
- Dennen, V. P. (2008). Looking for evidence of learning: Assessment and analysis methods for online discourse. *Computers in Human Behavior*, 24, 205-219. doi:10.1016/j.chb.2007.01.010
- Dini, D., & Liu, L. (2017). Intrinsic motivation factors in gamified photography learning: Direct and indirect effects. *Journal of Educational Technology Development and Exchange*, 10(2), 1-24.
- Dyment, J. E., & Downing, J. (2018a). Online initial teacher education students' perceptions of using web conferences to support professional conversations. *Australian Journal of Teacher Education*, 43(4), 68-91. doi:10.14221/ajte.2018v43n4.5
- Dyment, J., & Downing, J. (2018b). There was nowhere to hide: The surprising discovery of how weekly web conferences facilitated engagement for online initial teacher education students. *Asia-Pacific Journal of Teacher Education*, 46(4), 399-418, doi:10.1080/1359866X.2018.1444140
- Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*. Retrieved from <http://imej.wfu.edu/articles/1999/2/04/index.asp>.
- Gagne, R. M., Wager, Q. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design* (5th ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33, 441-467. doi:10.1177/1046878102238607.
- Google Trends. (2018). <https://trends.google.com/trends>
- Hayes, B., & Ringwood, J. V. (2008). Student authentication for oral assessment in distance learning programs. *IEEE Transactions on learning technologies*, 1(3), 165-175. doi:10.1109/TLT.2009.2
- Holzweiss, P. C., Joyner, S. A., Fuller, M. B., Henderson, S., & Young, R. (2014).

- Online graduate students' perceptions of best learning experiences. *Distance Education*, 35(3), 311-323. doi:10.1080/01587919.2015.955262
- Jaggars, S. S., & Xu, D. (2016). How do online course design features influence student performance? *Computers & Education*, 95, 270-284. <https://doi.org/10.1016/j.compedu.2016.01.014>
- Kent, C. Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education*, 97, 116-128. doi:10.1016/j.compedu.2016.03.002
- Klisc, C., McGill, T., & Hobbs, V. (2009). The effect of assessment on the outcomes of asynchronous online discussion as perceived by instructors. *Australasian Journal of Educational Technology*, 25(5), 666-682. Retrieved from <https://ajet.org.au/index.php/AJET/issue/view/28>
- Lee, A. M., & Liu, L. (2016). Examining flipped learning in sociology courses: A quasi-experimental design. *International Journal of Technology in Teaching and Learning*, 12(1), 47-64.
- Liu, L. (2015). Trends of research, practice and publishing: A comparison of content analysis in SITE Research Highlights. Panel presentation and discussion at SITE (Society of Information Technology in Education) 2014, Las Vegas, NV. March 4, 2015.
- Liu, L. (2018). A model of dynamic instructional design for dynamic learning and assessment. Keynote speaker at Research Forum of Southwest University, Chongqing, China, November 20, 2018.
- Liu, L., Chen, L., Li, W., Zou, C., & Pugh, K. (2018). Conducting web conferences to verify student identify and assess learning outcomes in online courses. Society for Information Technology & Teacher Education International Conference 2018 (Mar 26, 2018) pp. 208-213.
- Liu, L., & Gibson, D. (2018). A model of dynamic instructional design and dynamic assessment: Applied in teacher education online courses. Presented as a panel leader at Society for Information Technology & Teacher Education International Conference 2018 (Washington DC, Mar 26, 2018).
- Liu, L., Gibson, D., & Maddux, C. (2013). Writing to publish in SITE Research Highlight Book: Aids and cautions. Panel hosted at SITE (Society of Information Technology in Education) 2013, New Orleans, LA. March 25-29, 2013.
- Liu, L., & Velasquezbryant, N. J. (2003). An information technology integration system and its life cycle: What is missing? *Computers in the Schools*, 20 (1/2), 91-104.
- Liu, L., & Johnson, D. L. (2002). Assessing student learning in instructional technology courses within the dimensions of learning model: Static versus dynamic assessment. *Computers in the Schools*, 18(2/3), 79-95.
- Liu, L., Ripley, D., & Lee, A. (2016). Flipped learning and influential factors: Case analysis. *Journal of Educational Technology Development and Exchange*, 9(2), 89-107.
- Lowe, T. W. (2015). Online quizzes for distance learning of mathematics. *Teaching Mathematics and its Applications*, 34, 138-148. doi:10.1093/teamat/hrv009
- Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, 13(3). Retrieved from <https://ro.uow.edu.au/jutlp/vol13/iss3/7/>
- MAXQDA. (2014). The art of data analysis. Retrieved from <http://www.maxqda.com/products/maxqda10>. Berlin: VERBI GmbH.

- Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem solving. *Computers in Human Behavior*, 18, 623–632. doi: 10.1016/s0747-5632(02)00020-1.
- Mayotte, S. (2010) Online assessment of problem solving skills. *Computers in Human Behavior*, 26, 1253–1258. doi:10.1016/j.chb.2010.03.033.
- Mundkur, A., & Ellickson, C. (2012). Bringing the real world in: Reflection on building a virtual learning environment. *Journal of Geography in Higher Education*, 36(3), 369-384. doi: 10.1080/03098265.2012.692073
- Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal*, 89, 76–91. doi: 10.1111/j.0026-7902.2005.00266.x.
- Nyland, R., Davies, R., Chapman, J., & Allen, G. (2017). Transaction-level learning analytics in online authentic assessments. *Journal of Computing in Higher Education*, 29(2), 201-217. doi:10.1007/s12528-016-9122-0
- Okada, A., & Scott, P. (2015). Effective web videoconferencing for proctoring online oral exams: A case study at scale in Brazil. *Open Praxis*, 7(3), 227-242. doi:10.5944/openpraxis.7.3.215
- OWL. (2018). Meeting OWL. Available at <https://www.owllabs.com/>
- Scott, L. A., Temple, P., & Marshall, D. (2015). UDL in online college coursework: Insights of infusion and educator preparedness. *Journal of Asynchronous Learning Network*, 19(5), 99–120
- Uribe, S. N., & Vaughan, M. (2017). Facilitating student learning in distance education: A case study on the development and implementation of a multifaceted feedback system. *Distance Education*, 38(3), 288-301. doi:10.1080/01587919.2017.1369005
- Zlatović, M., Balaban, I., & Kermek, D. (2015). Using online assessments to stimulate learning strategies and achievement of learning goals. *Computers & Education*, 91, 32-45. doi: 10.1016/j.compedu.2015.09.012

#### **Contact the Author**

##### ***Leping Liu***

Professor,  
Information Technology and Statistics  
University of Nevada, Reno, USA.  
Email: liu@unr.edu

##### ***Li-Ting Chen***

Assistant Professor,  
Educational Measurement and Statistics  
University of Nevada, Reno, USA.  
Email: litingc@unr.edu