



# Exploring Lecture Evaluation Tools Suitable for Online Classes at Universities

Youngsun Song<sup>1</sup>, Jisook Yeom<sup>2</sup>

<sup>1</sup>Institute for High Education Innovation, Konkuk University, Chungju, Korea

<sup>2</sup>Department of Early Childhood Education, Konkuk University, Chungju, Korea

**Purpose:** To develop an evaluation tool suitable for online classes at universities due to the increased focus on the quality management of university classes because of the impact of COVID-19.

**Methods:** Nineteen items relevant to evaluation of the teaching and learning at university were extracted from a review of previous studies and the content analysis by experts was undertaken. Based on the context-input-process-product evaluation model, 16 items were selected from the analysis of learning evaluations of 1,000 students at the end of the second semester of 2021 at the researcher's workplace.

**Results:** First, following confirmatory exploratory factor analysis (CFA), 16 learning evaluation items were converted into a single factor related to learning and teaching. Second, as a result of confirmatory factor analysis (CFA) by the academic field, 16 learning evaluation items were converted into a single factor. Third, 16 learning evaluation items were found to have a significant positive effect on the learning evaluation score.

**Conclusion:** The learning evaluation tools developed through this study were demonstrated to be meaningful in that they can be applied at the individual educator and university level to improve the quality of lectures in online classes in the future.

**Keywords:** COVID-19; Online class; Learning evaluation; CIPP evaluation model; Online learning

## INTRODUCTION

### The need and purpose of the study

As most universities have realized the need for online classes, distant classes have been fully implemented (Kim, 2020; Do, 2020). Online classes require a shift in teaching-learning methods from teacher-centered to student-centered (Graham et al., 2001; Shelton, 2011). For student-centered classes, it is necessary to select a class method that provides video content, which utilizes a video platform such as Zoom, Google meets, or Webex. However, some educators who are accustomed to offline classes have faced significant difficulties in conducting online classes. These educators' difficulties require the creation of a learning environment suitable for online situations, analysis of students' needs, and change of teaching methods (Borch et al., 2020). As a result, the professional development of the teaching competency of teachers in online classes is critical (Thomas & Graham, 2019), and implies that learning evaluation tools that can check whether or not they are adapted to online classes should be developed.

This study started with the recognition that a learning evaluation tool suitable for the online classes environment is needed. The tool can be used to evaluate the effectiveness and efficiency of classes; it is the most common and effective mechanism among the methods used to improve

© 2022 International Society for Problem-Based Learning

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

the quality of lectures (Kim, 2017; Lee, 2001; Song, 2018; Suárez et al., 2022). Through learning evaluation, educators receive feedback on class quality from students who are the subject of classes, and through this, reflect on and improve their classes to strengthen teaching competency and improve classes quality (Song & Lee, 2020). Therefore, the university's learning evaluation system is designed to first ensure educators realize the difference between the meaning they want to convey to the student and how the student perceives it. Second evaluation provides an opportunity for the teacher to reflect on the classes, and third to recognize the teacher as an essential element in university learning and teaching processes. Feedback from students is significant in that it reminds us of the importance of competency in managing learning events.

Online classes are similar to classes at distant universities in that the educators and the students are separated by location. However, some characteristics distinguish these from classes at distant universities in-class preparation, class operation, interaction, and learning evaluation processes. Therefore, there is a limit to using the existing learning evaluation tools for distance learning (Song & Lee, 2020). That is, an in-depth discussion is needed on whether appropriate and timely interaction between educators and student occurs when the offline learning process is converted to an online learning process (Jung & Yoon, 2020; Kim & Cheon, 2020; Lee, 2020; Lee & Kim, 2020; Lee et al., 2020). In addition, since learning evaluation is conducted in online manner, the issue of whether the evaluation method is appropriate for the online teaching method and whether the evaluation standard is fair is raised. Accordingly, teachers have the right to receive appropriate evaluations for online classes, and universities must develop and provide learning evaluation tools that are appropriate and reliable.

The development of learning evaluation tools for online classes should be considered in relation to several contextual factors. Among them, the student's ability to use computers and the internet, online interaction, interface design, and systems operation and management are important factors (Park et al., 2006; Seo, 2002; Cheung, 1998). In addition, the complementary nature and characteristics of the academic discipline are key factors and should be reflected in evaluation criteria.

Therefore, the purpose of this study was to develop a learning evaluation tool that can fully reflect the university's online classes activities that arose due to the COVID-19 situation. To achieve this purpose, we explored the factors of learning evaluation of online classes based on the review of previous studies, extract evaluation items as factors based on the CIPP evaluation model, and verify the validity of the teaching evaluation tool. To this end, first, domestic literature relating to evaluation of online classes was re-

viewed. Second, the factors of learning evaluation were extracted and an expert's content analysis was undertaken. Third, confirmatory factor analysis was conducted to test the validity of the learning evaluation tool. Through results from these actions, we derived learning evaluation tools suitable for the acquisition of knowledge and processes that students want to acquire as learners. Ultimately, this study has the potential to contribute to improving students' satisfaction with the learning events available to them in the online environment.

### *Research Questions*

Research Question 1. What are the factors and questions for learning evaluation suitable for online classes at universities?

Research Question 2. Are the factors and evaluation items of learning evaluation valid?

### **Study Variables and Previous Studies**

#### *Purpose and use of university learning evaluation*

Teaching evaluation at universities began in the United States in the 1960s to meet the responsibility of university education and the needs of students' rights to optimal learning environments. In Korea, the idea officially discussed in the 1980s, and since the 1990s, it has been implemented at all universities from a student-centered educational perspective, university education competitiveness, and enhancement of teacher competency (Song, 2018; Lee, 2013). Learning evaluation was included in the evaluation tools of the 'University Comprehensive Evaluation and Accreditation System' implemented in 1994; it was established to improve the quality of lectures (Yang, 2014; Lee, 2001). The evaluation system was meaningful to the university in that it impacted individual teachers' efforts and interests and became part of the university learning and teaching systems (Song, 2018).

The purposes of evaluation are formative and summative (Song, 2018; Han et al., 2005; Braskamp, 1984). Formative evaluation enables early feedback on the worthiness of lectures and provides information for improvement. Summative evaluation focuses on the value, effectiveness, and efficiency of educational programs but also utilizes the feedback as decision-making information needed as evidence for teacher promotion, re-appointment, and guaranteed retirement age (Baek & Shin, 2008). The rationale behind evaluation details is as follows: First, that educators have appropriate qualifications, and that their professionalism and accountability are continuously improved. Second The student, who is both the subject of the class and the subject of learning evaluation, provides feedback on the classes. Third, it is used to make administrative decisions such as evaluation of educational achievements and promo-

tion, teaching support, and selection of excellent educators. Fourth, information for selecting a class (teaching) is provided when a student registers for a course. In summary, learning evaluation is conducted to improve the quality of lectures, measure the effectiveness of classes, provide basic data for selecting excellent teachers, and develop an excellent teaching model (Kim et al., 2007; Kim & Kim, 2008; Park, 2012).

Learning evaluation is a means of judging the value of a class activity (Braskamp, 1984) and an index to measure quality. Learning evaluation is carried out in the same context as the learning purpose (Baek & Shin, 2008). Educators are encouraged to reflect on their lectures and make an effort to improve them. The university examines teaching achievements or teaching excellence, providing incentives for improvement, and might apply sanctions where evaluation outcomes are poor. Therefore, the purpose and use of learning evaluation are not only for reinforcing teaching capacity and improving the quality of lectures, but also providing data for academic performance evaluation, providing information to support students' right to choose classes, collect information for advice on teaching activities, and research on various options for classes. In this study, the concept of learning evaluation is defined as the act of students judging the value of all teaching-learning activities according to a set standard subject against certain criteria and methods with the aim of increasing the effectiveness of the classes.

#### *Review of previous studies on learning evaluation*

Learning evaluation involves a system with tools to measure the quality of classes for feedback to educators and to secure the quality of university classes for the students' learning experience (Choi et al., 2018). Previous studies involving traditional teaching methods include the following. First, various domestic studies were examined, focusing on the validity and reliability of learning evaluation tools (Song, Ji, 1994; Yum, 2008; Lee et al., 2005; Aleamoni, 1981; Preece, 1990). Song and Ji (1994) reported that in one university, the learning evaluation tools were consistent with the purpose of the evaluation. By comparing the learning evaluation annually basis over two-years, they showed the ranking and measurement value of satisfaction were the same. Yum, 2008, and Heckert et al. (2006) cautioned that learning evaluation is a product of student, teacher, and student-teacher interaction and has multidimensional properties.

Looking at overseas studies, Aleamoni (1981) found that learning evaluation is a source from which important information can be obtained through various interactions with teachers within the learning environment. From a positive perspective teaching and learning can be evaluated logically, improving the quality of lec-

tures and providing opportunities for active participation around feedback. Also, Stewart et al. (2013), and Sun et al. (2008) reported that there is a correlation with learning satisfaction. However, Preece (1990) highlighted negative aspects noting that as an evaluator, the student lacks experience and professionalism, can focus on popularity rather than the evaluation of teaching ability. Concerns were raised about the reliability and validity of the evaluation item itself, and the consistency and stability of the item response. As such, it can be seen that both positive and negative opinions exist in the validity and reliability of learning evaluation.

Second, looking at previous studies on factors that affect learning evaluation, in Kim's (2005) study, students' learning motivation, expected grades, classes burden, and personal difficulties had an effect. The paper did not provide details on the number of students and teaching methods. In the study by Ting (2000), the higher the teacher's position, the more positive the learning evaluation. However, in the study by Coshins (1988), the teacher's age, years of service, gender, race, personality, and research performance did not affect learning evaluation. Han (2001) stated that teachers who taught humanities and social science subjects tended to receive higher scores than teachers who taught natural engineering subjects. In addition, it was said that the older the teacher, the lower the score, but the higher the age distribution and the more diverse the majors of the students, the higher the score. In research by Marsh (1984), the learning evaluation of small lectures was higher than that of large lectures, but in that by Han (2001), it was the opposite. In the study by Yum (2008), variables such as expectations for credits, lecture composition, and progress had a significant effect on evaluation, but the expectation around completion was not significant. Hence the factors affecting the learning evaluation were different for each study.

Third, the focus was on learning evaluation; according to studies by Han, Lee, Kim, (2005), and Kim et al. (2007), most universities only disclose learning evaluation data to individual educators, reflecting performance evaluation rather than class improvement. It was being used for administrative purposes, such as sanctions against teachers, that is those with low learning evaluation scores. In the study by Kim (2008), educators with high research achievements had higher learning evaluations than those with relatively low research achievements. Hence while studies on the use of learning evaluation are conducted for classes improvement, in reality, it can be seen that they are used for various administrative purposes rather than class improvement. These results suggest that clarity within discussion about learning evaluation is needed.

*Learning evaluation tool development process and use of CIPP evaluation model*

Using Stufflebeam's (1971, 2004) CIPP evaluation model, tools were developed for context, input, process, and product factors. The CIPP evaluation model is not limited to class content and calculates comprehensive class information by evaluating the situation within and resources invested in class operation, evaluation of resources, evaluation of the operating process, and the value of operating results. The advantage of the CIPP evaluation model is that it considers aspects of context, input, process, and product to make a comprehensive evaluation of the online classes scene.

The CIPP model highlights what we should do in online classes, how should we do it and whether we are doing it right, that is how well the approach to teaching was undertaken. Context (C) evaluates the worthiness or justifiable grounds for determining the class goal. The current situation and preferred method will have been defined. Input (I) centers on the information necessary for decisions on how to use resources to achieve the online class objective, selected methods and implementation plans. Process (P) evaluation, which checks the various methods required during the class and evaluates tools planned in advance, that is, the input. Product (P) evaluation, measures and interprets the achievement of class goals, both intended and unintended outcomes (Figure 1).

Evaluation of Context helps to establish the purpose of developing tools at the beginning stage of item development, and establishing the worthiness of the learning evaluation purpose on completion of the tool development. A focus on Input helps to identify differences and problems between the situation shown in class

against the intended situation. Product evaluation collects and evaluates a wide range of information related to the achievement of goals, thus helping to make decisions about improving learning evaluation. The reason why the CIPP evaluation model was used in this study is that it has the advantage of being able to evaluate learning events as needed at any stage.

**METHOD**

**Research Procedures**

To develop learning evaluation tools for online classes at universities, research procedures were carried out as shown in Figure 2 below. First, sub-factors and preliminary tools were derived by analyzing the literature review related to online classes at universities, learning evaluation tools in offline classes, and learning evaluation tools in distant classes. An expert review was conducted to secure the validity of the content of the preliminary tools. Five experts included for the expert panel were selected - two majors in education, one in educational engineering, and two in human resource development policy, with a doctorate.

Tools were corrected, items added, or deleted based on the calculated results. In addition, 30 students who had experience in online classes were involved in testing for face validity, and it was confirmed that there was no difficulty in understanding the items. Finally, the validity of the tools was verified by analyzing the data of 1,000 students in the mid-term and final learning evaluations conducted at K University located in a non-metropolitan area. The scale for each item response was measured on a 5-step Likert scale

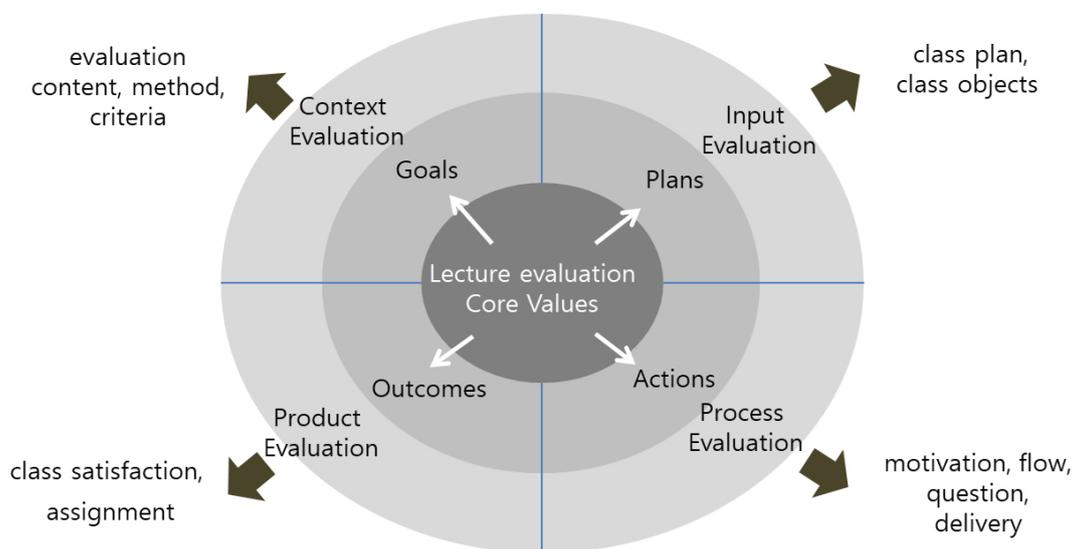


Figure 1. Factors for lecture evaluation using the CIPP evaluation model.

Sub-factor extraction and preliminary tools of the lecture evaluation tool Development	Sub-factor extraction and initial item development through literature analysis	
		<ul style="list-style-type: none"> <li>• Literature analysis related to lecture evaluation tools of existing university lectures and distance classes</li> <li>• Exploring online classes and developing initial items</li> </ul>
		⇩
	Verification of content validity and confirmation of preliminary items	
Feasibility study of a lecture evaluation tool		Investigate
		<ul style="list-style-type: none"> <li>• Analysis of lecture evaluation data for 1,000 students who conducted mid-term and final lecture evaluations</li> </ul>
		⇩
	Perform tool justification	
		<ul style="list-style-type: none"> <li>• Descriptive statistical analysis, exploratory factor analysis, regression analysis</li> </ul>

Figure 2. Lecture evaluation development process for online classes.

(not at all - strongly agree). In this study, descriptive statistical analysis, preliminary questions, exploratory factor analysis, and regression analysis of this questions were used.

The mean (M), standard deviation (SD), skewness, and kurtosis of each item were confirmed through descriptive statistical analysis. When the standard deviation is at least .15, the item is judged to be appropriate (Meir & Gati, 1981). For skewness, the absolute value is 3.0 or less, and for kurtosis, if the absolute value is 10.0 or less, the response data is judged to be normal (Kline, 2005). For exploratory factor analysis, the validity of the learning evaluation tools for online classes consisting of a screening test and cumulative division ratio review, a total of 1 factor, and 16 items were confirmed.

**Study participants**

Evaluations data from 1,000 university students from Year 4 at K-University located in a non-metropolitan area was used for this study. These were students who completed mid-term and final learning evaluations in 2021. The preliminary examination, the mid-term learning evaluation, is held from October 13 to 26, and the final evaluation, after the main examination, is from December 4 to 12. For the reliability of the study results, subjects with 30 or more students were extracted as a stratified sample. In addition, a

course with fewer than 30 students may give a good impression due to direct or frequent interaction with the teacher and maybe more positively viewed in learning evaluations. After dividing into 5 academic disciplines designated by the Ministry of Education, such as [Humanities and Social Sciences], [Natural Science], [Engineering], [Pharmaceuticals], and [Arts and Physical Education], they were extracted in similar discipline proportions to the student population. However, [Pharmaceuticals] academic discipline was excluded from the study as the number of potential cases was 21 (Table 1).

**RESULTS**

**Development of items for each factor in online classes learning evaluation**

To develop the tools, first, the factors and tools of the learning evaluation of the online classes were identified based on the analysis of previous studies related to learning evaluation. ‘class preparation’, ‘class operation’, ‘class output’, and ‘class evaluation’ factors were extracted by analyzing previous studies related to online classes conducted since the beginning of COVID-19. By applying this to the CIPP evaluation model, 4 factors of Context-Input-Pro-

**Table 1.** Characteristics of demographic variables

	Division	Frequency	Percentage
Gender	Male student	360	36
	Female student	640	63
Academic discipline	Humanities and social sciences	440	44
	Natural science	285	28.5
	Engineering	100	10
	Arts and physical education	175	17.5
	Grade	1 grade	280
Grade	2nd grade	249	24.9
	3rd year	311	31.1
	Grade 4	160	16
Sum		1000	100

cess-Product and 19 preliminary tools were developed as potential sub-factors of the online learning evaluation tool. Next, content validation was carried out through Delphi analysis by experts on a 5-point scale for preliminary tools.

First, the items involving the 4 factors above were seen as ‘too detailed’; these were revised. For example, the item is ‘Did the teacher provide information on classes activities for each week, classes changes, and assignments/exams promptly?’ After collecting the opinions of experts, it was revised to ‘Did the teacher compose the appropriate amount of learning to achieve the learning goal?’ In addition, ‘Teacher provided information about the online classes environment (e.g., LMS environment setting, pre-class procedures, bulletin board participation, etc.) promptly.’ Did you provide it in advance?’ and ‘In the syllabus, class information including class type (online, offline, mixed) and class method (e.g., real-time video lecture, provision of learning materials, provision of learning materials + real-time lecture, etc.) was specifically guided? was changed to ‘Did the teacher specifically present the class content and teaching method for one semester in the syllabus?’ Second, four items were excluded by reflecting the opinion that there is a possibility that they are similar or overlapping with other items. For example, ‘Is the amount of online learning activities (real-time video lectures, class materials, discussion/discussion, assignments, etc.) adequate?’; ‘Did the teacher deliver the content in an easy to understand way?’ The classes were conducted considering the level of the students,’ ‘I will recommend this class to my seniors and juniors and my classmates.’ In addition, for smooth communication with students in the online classes, the item ‘Did the teacher proficiently handle the various functions of Webex?’ was added to the ‘Input’ factor and the ‘Process’ factor, respectively. Finally, 4 factors and 16 items were confirmed: 4 items for Input, 7 items for

Process, 3 items for Product, and 2 for Context. It was confirmed that there was no difficulty in understanding the item, thus proceeding with the learning evaluation by conducting a face validity test on 30 university students who had experienced online classes (Table 2).

### Validation of learning evaluation tools for online class

#### *Descriptive statistics analysis result for each item*

Table 3 shows the results of descriptive statistical analysis of the final learning evaluation conducted for university students who participated in online classes to check the response level, distribution, and regularity of the learning evaluation tools in online classes. First, the average for each item was distributed in the range from 4.48 to 4.58, and the standard deviation was found to be in the range from .688 to .811. Next, as an index for confirming normality, the univariate skewness value and kurtosis value were derived, and the absolute value of skewness was 1.621-1.872, and the absolute value of kurtosis was 2.565-3.731. Therefore, it was found that there were no items that greatly deviated from the assumption of normality (Kline, 2005).

#### *Correlation of each item in the learning evaluation*

Examining the correlations for each item in the learning evaluation, there was found that there was a significant quantitative correlation at the level of .01 in all items. There was found that there was a significant quantitative correlation between the ‘learning evaluation score’ and the ‘learning evaluation items’ at the .01 level (Table 4).

#### *Verification of item validity and reliability*

Exploratory factor analysis (EFA) was performed to secure the validity of the items. The maximum likelihood analysis method was selected and conducted with a ‘direct oblimin’ factor rotation. As a result of KMO and Bartlett’s sphericity test to confirm the suitability of factor analysis, KMO = .982, which was found to be good (Kaiser, 1974). Bartlett’s sphericity is shown as  $X^2 = 474045.681$ ,  $df = 120$ , and  $Sig = .000$ , suggesting that it is worthwhile for factor analysis (Kang, 2013). The commonality was between .742 and .828, indicating that all items were important variables. To determine the number of factors for estimating the basic structure of all items, the scree test was first checked. Looking at the difference in the eigenvalues of the reduced correlation matrix as a result of the screening test, it can be seen that the eigenvalue decreases significantly from the 1st to the 2nd (difference in the eigenvalue 12.429), but does not decrease much (.122) and is normalized from the 3rd eigenvalue. Accordingly, it can be judged that

**Table 2.** Lecture Evaluation tools by CIPP

No.	CIPP	Tools	Questions
1	Input	Webex leverage	The teacher proficiently dealt with various functions of Webex.
2		Pre-classes information	The teacher informed me in advance whenever there was a change in classes.
3		Classes environment information provided in advance	The teacher provided information about online classes environments such as TLS and Webex in advance.
4		Presenting a specific lesson plan	In the syllabus, the teacher specifically presented the class's content and method for one semester.
5	Process	Classes proceed according to the syllabus	The teacher taught the classes according to the class's content and method presented in the syllabus.
6		Present class goals for each class	The teacher presented the lesson objectives for each class.
7		Appropriateness of online classes materials	The online classes materials provided by the teacher were appropriate for understanding the class content.
8		Class method appropriateness	The teacher used various teaching methods to understand the contents of the classes.
9		Efforts to interact with students	The teacher tried to actively interact with the students
10		Quick response to inquiries	The teacher quickly responded to the students' questions.
11		Summary of the main contents of the classes	Before the end of the classes, the teacher organized and delivered the main contents of the classes.
12	Product	Help to understand the content of the course	This class helped me to understand the content of the subject.
13		Overall satisfaction	I am overall satisfied with this class.
14	Context	Fidelity of evaluation contents	The test questions, assignment topics, and reports presented by the teacher fully reflected the content of the classes.
15		Clarity of evaluation criteria	The evaluation criteria for assignments, tests, and attendance suggested by the teacher were clear in the online classes.
16		Appropriate evaluation method	Online classes were fully considered for the evaluation methods of assignments, tests, and reports presented by the teacher.

CIPP: context-input-process-product.

**Table 3.** Mean and Standard Deviation of Lecture Evaluation tools

CIPP	Item	M	SD	Skewness	Kurtosis
Input	Webex leverage	4.50	0.770	-1.621	2.565
	Pre-classes information	4.57	0.714	-1.798	3.394
	Classes environment information provided in advance	4.57	0.700	-1.705	2.970
	Presenting a specific lesson plan	4.58	0.698	-1.747	3.199
Process	Classes proceed according to the syllabus	4.58	0.690	-1.739	3.123
	Present classes goals for each class	4.57	0.709	-1.743	3.080
	Appropriateness of online class materials	4.53	0.767	-1.758	3.162
	Class method appropriateness	4.52	0.766	-1.702	2.905
	Efforts to interact with students	4.55	0.752	-1.872	3.731
	Quick response to inquiries	4.56	0.734	-1.818	3.459
	Summary of the main contents of the classes	4.53	0.756	-1.691	2.843
Product	Help to understand the content of the course	4.52	0.774	-1.776	3.337
	Overall satisfaction	4.48	0.811	-1.695	2.845
Context	Fidelity of evaluation contents	4.58	0.688	-1.759	3.377
	Clarity of evaluation criteria	4.56	0.719	-1.727	3.057
	Appropriate evaluation method	4.56	0.721	-1.740	3.153

CIPP: context-Input-Process-Product.CIPP: context-input-process-product.

Table 4. Correlation between lecture evaluation items

Item	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰
① Webex leverage	1																
② Pre-classes information	.783**	1															
③ Classes environment information provided in advance	.804**	.852**	1														
④ Presenting a specific lesson plan	.778**	.822**	.847**	1													
⑤ Classes proceeds according to the syllabus	.775**	.813**	.837**	.882**	1												
⑥ Present class goals for each class	.768**	.800**	.819**	.850**	.861**	1											
⑦ Appropriateness of online class materials	.763**	.766**	.783**	.794**	.801**	.819**	1										
⑧ Class Method appropriateness	.781**	.775**	.792**	.798**	.800**	.814**	.851**	1									
⑨ Efforts to interact with students	.762**	.782**	.793**	.796**	.799**	.805**	.801**	.836**	1								
⑩ Quick response to inquiries	.763**	.783**	.793**	.796**	.797**	.803**	.784**	.806**	.855**	1							
⑪ Summary of the main contents of the class	.764**	.775**	.790**	.798**	.802**	.821**	.813**	.832**	.820**	.815**	1						
⑫ Fidelity of evaluation contents	.772**	.792**	.807**	.815**	.819**	.807**	.785**	.793**	.782**	.776**	.783**	1					
⑬ Clarity of evaluation criteria	.758**	.767**	.788**	.784**	.785**	.771**	.757**	.761**	.761**	.752**	.756**	.837**	1				
⑭ Appropriate evaluation method	.771**	.776**	.794**	.792**	.794**	.781**	.762**	.765**	.765**	.761**	.757**	.821**	.872**	1			
⑮ Help to understand the content of the course	.756**	.759**	.773**	.790**	.792**	.801**	.834**	.840**	.815**	.802**	.851**	.796**	.761**	.765**	1		
⑯ Overall satisfaction	.751**	.744**	.751**	.762**	.765**	.774**	.806**	.814**	.806**	.780**	.818**	.774**	.756**	.754**	.867**	1	
⑰ Lecture evaluation	.213**	.189**	.194**	.189**	.186**	.187**	.210**	.212**	.199**	.195**	.211**	.205**	.202**	.200**	.220**	.229**	1
	.859**	.877**	.893**	.905**	.904**	.899**	.890**	.899**	.892**	.888**	.892**	.882**	.873**	.870**	.895**	.880**	

CIPP: context-input-process-product  
 \*\* p < .01.

the structure of factors 1 and 2 immediately before leveling in the data matrix is an appropriate model. All items were extracted as one component (factor), and the total cumulative explanatory variance was 79.417%, which was found to be suitable as a factor model. If it shows more than 60%, it is recognized as a suitable model. The Cronbach  $\alpha$  value for 16 items conducted to check the degree of internal agreement was .984, which was found to be reliable.

After the screening test and review of the cumulative variance ratio, parallel line analysis was performed on the same data. In the parallel line analysis result, the eigenvalues obtained from both the sample and the wireless data were compared, and the number of eigenvalues showing a larger value in the sample data than in the wireless data is determined as the number of factors. In the case of the eigenvalue (12.917) derived from the actual data was higher than the eigenvalue (3.542) derived from the wireless data. However, from the case involving two factors, the eigenvalue (3.542) of the wireless data was higher than the eigenvalue (1.879) obtained from the actual data. In parallel line analysis, one factor was judged as the optimal factor because the upper limit of the number of factors was judged as the optimal number of factors within the range in which the eigenvalue of the actual data is larger than that of the wireless data. The components extracted by the maximum likelihood extraction method were found to be between .861 and .910 (Table 5).

Table 5. Initial components of lecture evaluation tools

CIPP	Items	Initial components
Input	Webex leverage	.861
Input	Pre-class information	.881
Input	Class environment information provided in advance	.900
Input	Presenting a specific lesson plan	.908
Process	Class proceeds according to the syllabus	.910
Process	Present class goals for each class	.908
Process	Appropriateness of online class materials	.892
Process	Class method appropriateness	.903
Process	Efforts to interact with students	.897
Process	Quick response to inquiries	.888
Process	Summary of the main contents of the class	.898
Product	Help to understand the content of the course	.898
Product	Overall satisfaction	.876
Context	Fidelity of evaluation contents	.893
Context	Clarity of evaluation criteria	.869
Context	Appropriate evaluation method	.874

CIPP: context-input-process-product.

The validity and reliability of the items by academic discipline were similar to the results extracted for all subjects regardless of academic discipline. The commonality of all items was higher than .60, indicating that all of them were important factors, and as a result of extracting the factors for each academic discipline, one factor was found, the range of factor loading was higher than .60, and the explanatory power was 75% or more (Table 6).

#### *Effect of learning evaluation tools on learning evaluation scores*

To investigate the effect of learning evaluation tools on learning evaluation scores, multiple regression analysis was performed using the learning evaluation tools as independent variables, learning evaluation score items as dependent variables, and demographic variables as control variables. As shown in Model 1, the educational background control variable was found to affect the learning evaluation score. In Model 2, it was found that the control variables, academic background, and learning evaluation tools, had a significant positive effect on the learning evaluation score. The regression model shows  $F = 380409.843$  ( $p < .01$ ), and  $R^2 = .997$  for the regression equation, showing 99.7% of explanatory power. The condition that the tolerance limit of multicollinearity should be greater than or equal to .1 and the coefficient of variance expansion (VIF) should be less than or equal to 10 was satisfied (Table 7).

## DISCUSSION AND IMPLICATION

As COVID-19 persists, many universities have switched to online classes. We are working to maintain the quality of lectures during to the transition to and continuation of online classes that are unfamiliar to both teachers and students. Accordingly, this study was intended to provide basic data for improving the quality of online classes and to contribute to strengthening teaching competency by developing learning evaluation tools that reflect the characteristics of online classes. A discussion of the results derived from this study is as follows.

First, it was confirmed that the CIPP evaluation model developed by [Stufflebeam \(2004\)](#) is suitable for developing learning evaluation tools for online classes. CIPP is an evaluation model developed and developed by Stufflebeam. Compared to the existing evaluation model, the CIPP evaluation model can cover a wider range of subjects as evaluation targets, and it is possible to evaluate various components of the program, and it is characterized by a systematic approach from decision-making to evaluation. The context evaluation of the CIPP evaluation model helps to establish the purpose of developing the class evaluation tool at the beginning stage. It helps to evaluate the purpose of class evaluation established after the completion of the development of the evaluation tool for online classes. Input evaluation is useful for developing or

**Table 6.** Commonality of Lecture Evaluation tools by Academic Discipline

CIPP factor	Items	Humanities and Social Sciences	Natural Science	Engineering	Arts and Physical Education
Input	Webex leverage	.841	.881	.867	.876
Input	Pre-class information	.875	.89	.887	.879
Input	Class environment information provided in advance	.891	.911	.897	.903
Input	Presenting a specific lesson plan	.904	.911	.917	.908
Process	Class proceeds according to the syllabus	.905	.912	.915	.914
Process	Present class goals for each class	.902	.916	.907	.908
Process	Appropriateness of online class materials	.879	.901	.883	.912
Process	Class method appropriateness	.891	.914	.894	.915
Process	Efforts to interact with students	.893	.898	.902	.899
Process	Quick response to inquiries	.888	.884	.882	.895
Process	Summary of the main contents of the class	.888	.912	.884	.905
Product	Help to understand the content of the course	.887	.912	.876	.910
Product	Overall satisfaction	.866	.888	.869	.879
Context	Fidelity of evaluation contents	.889	.889	.907	.904
Context	Clarity of evaluation criteria	.870	.871	.852	.871
Context	Appropriate evaluation method	.869	.887	.831	.884
Distributed explanatory power (%)		78.117	8.77	78.465	8.593

CIPP: context–input–process–product.

**Table 7.** Results of regression analysis between lecture evaluation tools and lecture evaluation scores

Independent variable	Model 1(control variable)				Model 2(independent variable)			
	B	SE	$\beta$	t	B	SE	$\beta$	t
(Constant)	90.536	.475		190.58	-.397	.046		-8.653
Grade	-.631	.088	-.050	-7.163**	.014	.005	.001	2.759**
Webex leverage					1.351	.013	.079	102.230**
Pre-class information					1.239	.016	.067	78.140**
Class environment information provided in advance					1.137	.018	.06	64.067**
Presenting a specific lesson plan					1.293	.019	.068	69.404**
Class proceeds according to the syllabus					1.27	.019	.066	66.818**
Present class goals for each class					1.245	.017	.067	71.211**
Appropriateness of online class materials					1.729	.015	.100	115.951**
Class method appropriateness					1.653	.016	.096	103.524**
Efforts to interact with students					1.566	.016	.089	98.101**
Quick response to inquiries					1.414	.016	.079	90.035**
Summary of the main contents of the class					1.931	.015	.111	127.759**
Help to understand the content of the course					1.284	.017	.075	75.810**
Overall satisfaction					1.442	.014	.089	99.835**
Fidelity of evaluation contents					1.502	.017	.078	88.315**
Clarity of evaluation criteria					1.376	.017	.075	82.494**
Appropriate evaluation method					1.364	.017	.074	82.210**
R <sup>2</sup> (.adj R <sup>2</sup> )		.007(0.006)					0.997	
$\Delta R^2(p)$		.007					.990	
F(p)		46.542 (p=.001)					380409.843 (p=.000)	

Control variables are treated as dummy variables

\*\* $p < .01$ , Dubbin-Watson=1.941.

evaluating the tools. Process evaluation helps to identify differences and problems between the situation shown in the course of the classes and the intended situation. Product evaluation collects and allows for judgment about a wide range of information related to the achievement of goals, thus helping to make decisions about improving class evaluation. The reason why the CIPP evaluation model was used in this study is that it has the advantage of being able to be used to evaluate any stage in the course of class implementation.

Second, 1 factor and 16 items were derived for learning evaluation of online classes through the analysis of previous studies on learning evaluation tools and previous studies related to online classes at universities due to COVID-19. By the procedures of class preparation, class operation, class product, and learning evaluation, the main items of learning evaluation were derived. Specifically, the input evaluation consisted of four items, including 'ability to use Webex', 'pre-class guide', 'preparation of class environment information', and 'presentation of a specific lesson plan'. The process evaluation consists of 7 items: 'Proceeding according to the syllabus', 'Suggesting class goals for each class', 'Appropriateness of online class materials', 'Appropriateness of class method', 'Efforts to

interact with students', 'quick response about questions' and 'organization of the core contents of the class' were included. The product evaluation consisted of two items, including 'help in understanding subject content' and 'overall satisfaction'. Finally, the Context evaluation included three items, 'fidelity of evaluation content', 'clarity of evaluation criteria', and 'appropriateness of evaluation method'.

Third, whereas the previous offline classes evaluation questions of universities were composed mainly of the teacher's class behavior, the revised tools were structured to reflect students studying independently. For example, learning support, that is, providing an environment for video classes, platform operation ability such as Webex, preparing materials for a video class, providing a class guide by week, and evaluation criteria, contents, and methods suitable for video classes are included.

Fourth, validation of the learning evaluation tools for online classes at universities was conducted based on the mid-term and final learning evaluation tools used for online subjects offered in the second semester of the 2021 academic year. As a result of exploratory factor analysis, it was found that 16 learning evaluation items were grouped into one factor in the scree test and cumulative vari-

ance ratio. It could be judged that all the items for each CIPP factor fully incorporated online learning events in universities. In this study, the fact that learning evaluation tools in online classes were not subdivided into different factors according to the classes procedure or teacher's behavior but appeared as one factor was linked to classes preparation, classes operation, classes assignments, and notebook evaluation in university online classes. This means that they are performed simultaneously; all CIPP factors should be considered together for quality control of online classes.

Fifth, the learning evaluation factors and items for online classes at universities presented in this study provide basic data for systematic teaching method development or teaching support to enhance classes satisfaction. You will be able to understand the difficulties faced by educators and students in online classes and ways to provide support.

This study explored learning and teaching factors that can be used to develop learning evaluation tools for online classes at universities and then verified the validity of the items through expert content analysis and exploratory factor analysis. Suggestions for future research are as follows. First, it is necessary to develop learning evaluation tools by including a series of processes such as determining convergent validity, criterion validity, and construct validity. Second, various learning evaluation methods should be reviewed. It is necessary to include use a mobile platform in consideration of the characteristics of students who are the smart generation as well as the online method implemented by most universities. In addition, various descriptive items should be developed to evaluate lectures quantitatively as well as qualitatively so that they can be referred to for lecture improvement.

While this study provides exploratory level result, it provides basic data for developing and validating learning evaluation tools for various online classes by reflecting the future academic disciplines and teaching methods such as theory, practical skills, practice, and experiments.

## CONFLICT OF INTEREST

The authors declared no conflict of interest.

## ACKNOWLEDGEMENTS

This paper was supported by funding from the Halla Newcastle PBL Education and Research Center.

## REFERENCES

- Aleamoni, L. M. (1981). Student ratings of instruction. *In handbook of teacher education*, edited by Jason Millamn. Beverly Hills, CA: Sage Publication.
- Baek, S. G., & Shin, H. J. (2008). Multilevel analysis of the effects of student and course characteristics on student course evaluation-focused on the undergraduate liberal education program-. *Education Evaluation Research*, 21(2), 1–14.
- Bojović, Ž., Bojović, P. D., Vujošević, D., & Šuh, J. (2020). Education in times of crisis: rapid transition to distance learning. *Computer Applications in Engineering Education*, 28(6), 1–23.
- Borch, I., Sandvoll, R., & Risør, T. (2020). Discrepancies in purposes of student course evaluations: what does it mean to be “satisfied”? *Educational Assessment, Evaluation and Accountability*, 32(1), 83–102.
- Braskamp, L. A. (1984). *Evaluating teaching effectiveness: A practical guide*. Sage.
- Cheung, D. (1998). Developing a student evaluation instrument for distance teaching. *Distance Education*, 19(1), 23–42.
- Chisadza, C., Nicholls, N., & Yitbarek, E. (2019). Race and gender biases in student evaluations of teachers. *Economics Letters*, 179, 66–71.
- Choi, K. J., Kwon, S. Y., Kim, E. J., & Park, I. S. (2018). A study on the professors' perception of the course evaluation: focusing on a university. *The Journal of Learner-Centered Curriculum and Instruction*, 18(3), 415–437.
- Coshins, W. E. (1988). Student ratings of teaching: a summary of the research. Idea Paper, 20, Central for Faculty Evaluation and Development. Manhattan. Kansas: Kansas State University.
- Do, J. W. (2020). An investigation of design constraints in the process of converting face-to-face course into online course. *Journal of Education & Culture*, 26(2), 153–173.
- Graham, C., Cagiltay, K., Lim, B. R., Craner, J., & Duffy, T. M. (2001). Seven principles of effective teaching: a practical lens for evaluating online courses. *The technology source*, 30(5), 50.
- Han, S. I. (2001). Analysis of factors related to lecture evaluation by students: focusing on factors related to professors, students and class. *The Journal of Educational Administration*, 9(4), 247–266.
- Han, S. I., Lee, J. Y., & Kim, H. J. (2005). A comprehensive study of Korean students' evaluations of university teachings. *The Journal of Educational Administration*, 23(3), 379–403.
- Heckert, T. M., Latier, A., Ringwald Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to “buy” better evaluations through lenient grading? *College*

- Student Journal, 40(3), 588–597.
- Jung, H. Y., & Yoon, J. W. (2020). A survey research of student's perception of Korean language online video lecture. *The Journal of Humanites and Social science*, 1(3), 1305–1318.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
- Kang, H. (2013). A guide on the use of factor analysis in the assessment of construct validity. *Journal of Korean Academy of Nursing*, 43(5), 587–594.
- Kim, H. I., Kim, S. S., Kwon, O. Y., Le, C., & Row, K. H. (2007). Improving course evaluation system of engineering education. *Journal of Engineering Education Research*, 10(4), 58–77.
- Kim, J. H., Park, J. K., & Hong, Y. E. (2010). The study of how influences in the environment of college education affects students' lecture evaluation: based on data of Y university. *Korean Association of Business Education* (pp. 1–16).
- Kim, J. K. (2017). Developing a student evaluation instrument for college teaching. *Journal of the Korea Academia-Industrial Cooperation Society*, 18(6), 187–196.
- Kim, M. H. (2005). Validity and reliability of lecture evaluation. *Asian Journal of Education*, 6(3), 1–24.
- Kim, S. H., & Cheon, S. M. (2020). A case study of online class operation and instructor's difficulties in physical education as a liberal arts in university due to COVID-19. *Korean Society Of Sport And Leisure Studies*, 81(1), 9–26.
- Kim, S. M. (2020). Analysis of press articles in Korean media on online education related to COVID-19. *Journal of Digital Contents Society*, 21(6), 1091–1100.
- Kim, S. S., & Kim, H. I. (2008). Developing and validating midsemester student's evaluations of college teaching. *Journal of Educational Evaluation*, 21(1), 55–78.
- Kline, R. B. (2005). Principles and practice of structural equation modeling (2nd ed.). The Guilford Press.
- Lee, B. K. (2020). A study on learners' response to online college english class as general education due to the COVID-19 pandemic. *Korean Journal of General Education*, 14(4), 97–12.
- Lee H. W., Kang, H. S., Jung, Y. S., & Heo, Y. E. (2005). Analysis on Evaluation Inquiry of Lectures for the Improvement on University Lecture Quality: Focused on Liberal Art Courses of Engineering and Science Schools at SNU. *Journal of engineering education research*, 8(4), 52–63.
- Lee, D. J., & Kim, M. S. (2020). University students' perceptions on the practices of online learning in the COVID-19 situation and future directions. *Multimedia-Assisted Language Learning*, 32(3), 359–37.
- Lee, H. S. (2020). A study on the perception of professors and learners on the remote learning of university education: focused on the cases of M university. *Journal of the Korean School Mathematics Society*, 23(3), 37–395.
- Lee, S. H. (2001). A study on the present situation and problems of students' evaluation. *Journal of Educational Technology*, 17(1), 81–106.
- Lee, J. G. (2013). A study on the improvement of a lecture evaluation questionnaire in the university. *Korean Journal of General Education*, 7(6), 247–274.
- Lee, Y. H., Park, Y. J., & Yun, J. H. (2020). Exploring the "Types" through case analysis on operation of distance education in universities responding to COVID-19. *The Journal of Yeolin Education*, 28(3), 211–234.
- Marsh, H. W. (1984). Students' evaluation of university teaching: dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707–754.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In *The scholarship of teaching and learning in higher education: An evidence based perspective*. Springer, Dordrecht.
- Meir, E. I., & Gati, I. (1981). Guidelines for item selection in inventories yielding score profiles. *Educational and Psychological Measurement*, 41(4), 101–1016.
- Minister of Education (2020). Online education guidelines. Ministry of Education press release.
- Park, H. J., Choi, M. S., & Lee, G. M. (2006). Development and validation of a cyber course evaluation tool. *Journal of Education Evaluation*, 19(2), 203.
- Park, H. R. (2012). A study on the improvement of a lecture evaluation tool in higher education: A case of improvement of a lecture evaluation questionnaire in "A" university. *Journal of the Korea Academia-Industrial Cooperation Society*, 13(1), 5033-5044.
- Preece, M. (1990). The reliability, validity, and usefulness of student evaluations for the purpose of improving teaching in postsecondary institutions: Annotated bibliography. ERIC Clearinghouse.
- Schutt, Allen & Laumakis. (2009). The effects of instructor immediacy behaviors in online learning environments. *The Quarterly Review of Distance Education*, 10(2), 135–148.
- Seo, E. H. (2002). Development of students' evaluation instrument for cyber instruction. *Yonsei Review of Educational Research*, 15(1), 133–153.
- Shelton, K. (2011). A review of paradigms for evaluating the quality of online education programs. *Online Journal of Distance Learning Administration*, 4(1), 1–11.
- Song, H. D., & Lee, Y. C. (2020). Exploring factors and indicators for measuring quality of non-contact university lectures. *Global*

- Creative Leader: *Education & Learning*, 10(4), 245–273.
- Song, M. S., & Ji, E. L. (1994). An item analysis study on the lecture evaluation questionnaire. *Education Evaluation Research*, 7(2), 263–283.
- Song, Y. S. (2018). Analysis of factors affecting satisfaction of teaching evaluation in the university. *Journal of Competency Development & Learning*, 13(2), 139–161.
- Song, Y. S. (2021). Analyzing the mediating effect of learning styles in the relationship of the influence of project class methods on learning satisfaction. *Journal of Learner-Centered Curriculum and Instruction*, 21(1), 1453–1474.
- Stewart, B. L., Goodson, C. E., Miertschin, S. L., Norwood, M. L., & Ezell, S. (2013). Online student support services: A case based on quality frameworks. *Journal of Online Learning and Teaching*, 9(2), 290–303.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education*, 5(1), 19–25.
- Stufflebeam, D. L. (2004). *The 21st-century CIPP model: Origins, development, and use*. Thousand Oaks, CA: Sage.
- Suárez Monzón, N., Gómez Suárez, V., & Lara Paredes, D. G. (2022). Is my opinion important in evaluating lecturers? Students' perceptions of student evaluations of teaching (SET) and their relationship to SET scores. *Educational Research and Evaluation* (pp. 1–24).
- Sun, P. C., Tsai, R. J., Finger, G., Chen, Y., & Yeh, D. (2008). What drivers a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50(4), 1183–1202.
- Thomas, J. E., & Graham, C. R. (2019). Online teaching competencies in observational rubrics: What are institutions evaluating? *Distance Education*, 40(1), 114–132.
- Ting, K. (2000). A multi-level perspective on student ratings of instruction: lessons from the Chinese Experience. *Research in Higher Education*, 41(5), 637–661.
- Yang, K. S. (2014). Meta-research on the influence of course, instructor, and student characteristic in student evaluation of teaching at universities. *The Korean Journal of Educational Methodology Studies*, 26(2), 293–332.
- Yum, S. C. (2008). Validating students' ratings of teaching scale and analyzing multilevel models. *Education Evaluation Research*, 21(2), 25–52.