IJOLAE
Indonesian Journal on Learning and Advanced Education
http://journals.ums.ac.id/index.php/ijolae

A CIPP Model Evaluation: Advancing Learning Through Interest Based-Learning Media Program in Higher Education

Almubarak[1], Nahadi[2]✉, Siti Sriyati[3], Aidil Adhani[4], Sukmawati Yasim[5]
[1-3]Faculty of Mathematics and Science Education, University of Pendidikan Indonesia, Indonesia
[1]Faculty of Teacher Training and Education, University of Lambung Mangkurat, Indonesia
[3]Faculty of Education, Monash University, Australia
[4]Faculty of Modern Languages and Applied Linguistics, University of Limerick, Ireland

DOI: 10.23917/ijolae.v8i2.16045
Received: January 30th, 2026. Revised: April 15th, 2026. Accepted: April 23rd, 2026
Available Online: April 27th, 2026. Published Regularly: May, 2026

Abstract

Evaluation of interest-based education programs is an essential component of assessing a program's effectiveness in higher education. However, valid, contextual, and relevant instruments for curriculum development objectives, such as those in chemistry education, remain limited. The purpose of this study was to determine how to (1) design and develop program evaluation instruments based on the Context, Input, Process, Product (CIPP) model, (2) program effectiveness by looking at evaluation results in each aspect of the model, and (3) relevance between evaluation results and the Model structure. The study used an evaluative approach with the CIPP evaluation model as the main framework. The instrument was developed independently and validated by experts with a logical-theoretical-empirical approach, with samples of chemistry education students in the chemistry learning media lecture program. Data were collected using a questionnaire, analysed using a statistical approach, and the evaluation results were reported. The evaluation results show that (1) the designed and developed evaluation instrument meets the valid criteria (0.91 or high validity), (2) this program has high effectiveness with an average score of Context (3.75), Input (3.5), Process (3.0), and Product (3.5), and (3) the evaluation results show relevance to the model structure and can be used effectively. These results confirm that the CIPP model can not only measure program effectiveness, but also identify areas that still need improvement for continuous improvement. CIPP model-based evaluation can be an effective tool for assessing the effectiveness of interest-based lecture programs in higher education.

Keywords : CIPP model evaluation, educational assessment, interest-based learning, student learning outcomes, advanced learning, innovation learning, lifelong learning.

Corresponding Author:
Nahadi, Faculty of Mathematics and Science Education, University of Pendidikan Indonesia, Indonesia
Email: nahadi@upi.edu

1. Introduction

Program evaluation is vital to ensure the effectiveness of program implementation (Fitzpatrick et al., 2011). The evaluation results confirm the impact of program implementation on participants' or users' development. The correct evaluation model has a significant effect on the program being evaluated, even though there are unexpected findings (negative findings). By design, program evaluation includes many aspects to determine whether the program being assessed meets the requirements. The elements in question are the program itself, objectives, design, samples, costs, data types, analysis, reporting formats, reporting schedules, reports, and the use of evaluation results (Fitzpatrick et al., 2011). With a comprehensive evaluation model and robust measurement capabilities, the evaluation results are crucial findings and have a significant impact on the program being evaluated. Thus, program evaluation is a strategic step to determine a program's effectiveness for participants, users, leaders, and policymakers, and to identify ways to improve human quality through program implementation.

In education, the evaluation of educational programs not only measures the program's effectiveness but also serves as a basis for increasing motivation and retention of participants in teaching profession programs (Ávila & Fernández, 2025). The decision to become a teacher may be influenced by the extent to which the program affects them. This evaluation can be easily adapted to program evaluations in school environments (Martin & Carey, 2012). In addition, the evaluation results can provide educational leaders and policymakers with insight into how to develop effective programs that prepare prospective teachers for the needs of the times (Yang et al., 2021). In addition to providing insight, evaluating programs also shows the extent to which the evaluation design is flexible and relevant to measurement (Proger, 1971). Flexibility is one of the main factors in the educational evaluation process (Proger, 1971). In his writings, Proger (1971) also emphasised the importance of criticising the evaluation methods used to identify the most appropriate, flexible evaluation method for measuring a program.

The "context, input, process, product" (CIPP) model is among the most structured and holistic in program evaluation (Fitzpatrick et al., 2011). In their study, Fitzpatrick et al. (2011) explained that the CIPP model comprises interrelated components and process cycles. For example, the "context" aspect provides an overview of the needs that must be met, while fulfilment can be achieved through effective planning through the "input" aspect. The implementation of the program plan falls under the "process" aspect, and the description of the results obtained falls under the "product" aspect. Continuous improvement, flexibility, representativeness, theoretical evidence, and practice make CIPP the right model for evaluating programs. CIPP is not only a model for measuring an education program's effectiveness but can also identify gaps in evaluation results and serve as a basis for program improvement. These improvements certainly affect the program's quality, including user experience and program development. Therefore, an effective program can only be achieved through a program evaluation model that is relevant and appropriate for measurement (Fitzpatrick et al., 2011).

The CIPP model has been widely used and adopted by various studies. Its flexible and structured nature makes this model a strategic tool for assessing the quality of an educational program. For example, Alquraan et al. (2025) used the CIPP model to evaluate teacher education programs. In this research, a quantitative approach is used to determine the structural relationships among CIPP components based on the perceptions of prospective teacher students (Alquraan et al., 2025). The results show no direct relationship among the four CIPP structures (model 1), whereas model 2 shows that three CIPP components are related to the product components. Alquraan et al.'s (2025) recommendation is for further studies on samples at various universities and valid instruments that measure CIPP from the perspective of program participants. The results of this evaluation show that using CIPP is effective in identifying multiple program outcomes (Alquraan et al., 2025).

Using a different technique, Akhtar et al. (2024) integrated the International

Education Standards (IES) with the CIPP model to assess the effectiveness of business education programs at universities in India in meeting ethical standards. The integration of CIPP and IES indicates that the program aligns with IES 4 (Akhtar et al., 2024). This evaluation also captures students' perceptions of the program's implementation and the teaching methods used. In terms of results, this research provides insight into CIPP, yielding accurate information about the quality of business education programs facing ethical challenges. As in the study by Alquraan et al. (2025), the evaluation process involved only students, making generalisation difficult (Akhtar et al., 2024). These findings show that the CIPP model offers many insights and recommendations for its use as an evaluation tool. In addition, the CIPP model was also used to evaluate the teaching process of a nursing education program (Zhao et al., 2024). The evaluation produced many findings, including recommendations for program improvements.

Elsewhere, Irene (2023) used the CIPP model to evaluate the teacher education curriculum in a Philippine college environment. The CIPP model identified several findings, including inconsistencies in the curriculum, background analysis, needs, teaching strategies, and more. The results indicate that consistent and effective program evaluation is needed to present curriculum updates that meet participants' needs (Irene, 2023). In the same study, the CIPP model was also an effective evaluation tool for tourism education programs (Tuna & Başdal, 2021). Improving program quality is the main finding of the CIPP approach, including the detection of deficiencies and weaknesses in the curriculum's implementation. Although effective in its use, sample limitations and other findings must be interpreted carefully (Tuna & Başdal, 2021). With the CIPP model, Meiklejohn et al. (2023) also found that the implemented program improved collaborative skills in health education. The integration of a social-constructivist perspective into the CIPP process provides a clear picture of the program's effectiveness. The evaluation results show that the research is relevant to health education, curriculum developers, and higher education stakeholders seeking to implement interprofessional education in a structured and systematic manner (Meiklejohn et al., 2023).

The research described is highly relevant to the nature of the CIPP model, in which evaluation results can serve as a basis for future program improvement (Fitzpatrick et al., 2011). The research also further confirms that the use of CIPP is very representative for use as an evaluation tool in various educational programs such as health education (Meiklejohn et al., 2023), business education (Akhtar et al., 2024), tourism education (Tuna & Başdal, 2021), guidance and counseling (Martin & Carey, 2012), and teacher education including curriculum (Irene, 2023). The effectiveness of CIPP also confirms its usefulness for evaluating programs and curricula across various fields, broadening the scope of this model in the context of modern education (Dizon, 2023). Modernisation and digital culture require today's generation to adapt to exponential changes, particularly in developing 21st-century skills. In this context, higher education must not only focus on skill acquisition but also promote advanced learning, which emphasizes higher-order thinking, metacognitive awareness, and adaptive expertise. Furthermore, the integration of innovation learning becomes essential to foster creativity, problem-solving, and the ability to generate novel solutions in dynamic environments. Equally important is the cultivation of lifelong learning, ensuring that prospective teachers continuously update their competencies and remain relevant in an ever-evolving knowledge landscape (Kain et al., 2024). Program evaluation can be a strategic step in identifying and preparing student teachers to address the various challenges of the digital era. The results of the program evaluation can serve as a basis for lecturers, researchers, and policymakers to design and develop high-quality, sustainable education programs that produce prospective teachers who are technologically proficient, competent, superior, innovative, and critical (Wisudawati et al., 2022).

The program evaluated in this context is a chemistry learning program in higher education. This program focuses on how students understand chemistry content while utilising technology by using the results of the demand analysis at the beginning of learning. The lecture scenario in this program is also based on the results of the student interest analysis (Barke et al., 2012). In other words, the analysis of needs (interests) at the beginning of learning is the basis for what students must do during the program. This program is considered capable of improving the structures and mental models of prospective teachers, as Barke et al. (2012) explain regarding the influence of interest in science learning.

The use of interest in chemistry learning media lecture programs represents a form of innovation learning, where student engagement becomes a driver for creative knowledge construction (van Dinther et al., 2023). This approach also aligns with the principles of advanced learning, as it encourages deeper conceptual understanding and the development of higher-order cognitive skills. Moreover, by fostering intrinsic motivation and sustained curiosity, the program contributes to the development of lifelong learning dispositions, which are essential for continuous professional growth. The involvement of interest in understanding not only affects cognitive development but also strengthens participants' perceptions, motivation, and behaviour (Barke et al., 2009). Barke et al. (2012) emphasised that interest enables the construction of scientific mental structures and models, which can be integrated with current technology (Kwarikunda et al., 2020, 2021). However, the systematic and comprehensive evaluation of the effectiveness of implementing this program remains limited, especially in higher education settings.

The CIPP (Context, Input, Process, Product) evaluation model offers a comprehensive framework for evaluating the effectiveness of educational programs. The use of CIPP is not only an evaluation of the final results but also of the context and process behind the program's implementation (Prameswari et al., 2024). Previous studies and theories support the logical sequence that makes CIPP the appropriate evaluation model for measuring program effectiveness. Although it has been widely used as an evaluation tool across various educational programs, studies specifically evaluating chemistry lecture programs that use this model remain limited (Espinosa et al., 2013).

This study not only adapts the CIPP model as an evaluation framework but also as a foundation for developing a systematic evaluation instrument with a representative set of indicators and an evaluation tool for assessing program effectiveness (Vlachopoulos & Makri, 2024). With a CIPP-based approach, this study provides a robust evaluation tool and comprehensive information on program effectiveness, which has previously been rarely done in chemistry education. In addition, the results of this study can be a fundamental factor in decision-making for teachers, researchers, curriculum developers, and decision-makers in designing programs that prioritise student needs, especially in higher education (Ødegård et al., 2025). This study is expected to expand the scope of the CIPP model in educational evaluation research while offering new perspectives on integrating evaluation models with interest-based approaches to learning.

The study questions are, (1) how is the design and development of CIPP-based program evaluation instruments; (2) how is the effectiveness of the program assessed by looking at the evaluation results in each aspect of the CIPP model; (3) how is the relevance between the evaluation results and the structure of the CIPP Model.

2. Method

To answer the study questions regarding the design and development of evaluation instruments based on the CIPP model, the extent to which the effectiveness of the program is assessed with the instrument developed with CIPP, and the relevance of the evaluation results to the CIPP evaluation model structure, this study uses an evaluative approach with the CIPP model framework (Context, Input, Process, Product) (Fitzpatrick et al., 2011). With its flexible and comprehensive framework, this model can provide concrete evaluation guidelines for programs, starting with the implementation context, the inputs used, the learning process, and the products achieved (Harefa et al., 2020). This study includes several stages so that it is easy to understand and replicate as an element of sustainability and meaningfulness in the future:

a. Preliminary study

The initial stage is to conduct interest-based chemistry learning. This interest-based lecture program has been developed and implemented over the last 2 years in the chemistry learning media course in the chemistry education program. The course was chosen because it was designed to foster technological skills, scientific knowledge, and attitudes among prospective chemistry teachers. The development of this program has been aligned with the curriculum structure, including the study program's vision. After several years have passed, the lecture program needs to be evaluated macro, as previously the program assessment was only micro, such as learning outcomes, task progress, and other supporting factors (Worthington & Whittaker, 2006).

The program's effectiveness needs to be evaluated comprehensively using structured, relevant instruments. Microprogram assessment is insufficient to determine the program's effectiveness and depth of impact on students, which is the main reason for the need for an evaluation. This assessment also assesses how the program will continue in the future and the advantages and obstacles of implementing it. To address this condition, an appropriate instrument is needed to comprehensively evaluate the program's effectiveness.

b. Instrument design and development

After obtaining the results of the micro-evaluation of the program implementation, a literature review was conducted on interest factors, chemistry learning trends, and evaluation models relevant to the internal observations conducted previously. The Context, Input, Process, Product (CIPP) model is the evaluation framework because it can be applied in a structured, systematic manner to evaluate a program. In addition, the CIPP model is suitable for identifying the program's strengths and weaknesses, thereby enabling the program's sustainability to be assessed from the evaluation results. Therefore, the CIPP framework is the basis for creating program evaluation instruments.

The instruments used were designed and developed independently by involving several experts. The "logical-theoretical-empirical" approach is used (Friedenberg, 1995; Haynes et al., 1995). Logical feasibility refers to the instrument's feasibility test, measured using a validation questionnaire and expert assessment (expert validity). Aspects of expert assessment include format, language, content, and sustainability. Each indicator in the instrument is designed according to the CIPP model, which includes questions that assess the program's effectiveness.

Theoretical feasibility refers to mapping and adjusting the instrument indicators to the context of chemistry education, interests, and their relevance to evaluation in higher education (Lin et al., 2013). The primary source of instrument content references uses several references, especially in question development (Lee et al., 2019; Zhang et al., 2011). This mapping is intended to ensure that the instrument's construct aligns with the scientific context of chemistry education and the CIPP evaluation framework.

Empirical feasibility refers to how the results of the expert validation analysis using Aiken V to obtain a complete picture of the instrument validation (Retnawati, 2016). Feasibility is assessed in several categories: low (x < 0.4), medium (0.8 > x > 0.4), and high (x> 0.8) (Retnawati, 2016). The validators involved are lecturers with expertise in chemistry education, educational research and evaluation, and language education. If content validation is to be quantified, researchers can ask experts to fill out the validation assessment sheet. Based on the three experts' entries, the expert agreement index (validator agreement) is then calculated using the Aiken index.

$$V = \frac{\sum s}{n(c-1)}$$

V = rater agreement index regarding item validity, s is the score set by each rater minus the lowest score in the category used (s = r - lo), where r - the rater's chosen category score and lo is the lowest score in the scoring category, n is the number of raters, and c is the number of categories that can be selected by the rater (Yang et al., 2021).

c. Evaluation Preparation and Planning

After the instrument is declared suitable for use, the preparation and planning of the evaluation begin. The stages include determining the evaluation objectives, which not only assess program effectiveness but also examine how the program supports advanced learning outcomes, promotes innovation learning practices, and fosters lifelong learning competencies among students. The program being evaluated is the interest-focused learning media lecture program. The samples include second-year chemistry education students and those who have enrolled in the chemistry learning media course.

d. Data collection

After the objectives and instruments are determined to be suitable, the program is evaluated by the validator. Evaluators are selected based on several criteria. First, experts in educational assessment and evaluation should ensure that the appraisal is original. Second, the evaluator has researched assessment, evaluation, and instrument development. Third, their scientific background is in chemistry education, science education, or general education. The evaluator is also actively involved to ensure that the evaluation results can be used as a basis for improving the program's quality in the future. Data were collected using a Likert-based questionnaire, which was categorised into four aspects of the CIPP model (Context, Input, Process, Product) to assess effectiveness at the macro level.

e. Data Analysis

Data analysis used a statistical approach based on the results of Likert data recapitulation, with criteria such as 4 = very relevant/very appropriate/very helpful/etc., 3 = relevant/appropriate/helpful, etc., 2 = less relevant/quite relevant/less appropriate/less helpfuletc., 1 = not relevant/not appropriate/not beneficial/etc (Retnawati, 2016).

The determination of the criteria range uses a mathematical approach, drawing on several analogies (Kyriazos & Stalikas, 2018; Wright & Masters, 1982). First, the lowest Likert value is 1, and the highest is 4, so the total range is 4 - 1 = 3. Second, the interval can be calculated by dividing the total range by the number of categories. Third, 0.75 is used as the range between scales, yielding the criteria shown below. Fourth, the average score is calculated as the average of each aspect, as described below the interval formula.

- $Internal = \frac{Total\ Range}{Number\ Categories} = \frac{4-1}{4} = 0.75$
- $Average\ Scope = \frac{\sum Total\ Average\ Score\ of\ Dimensions}{Number\ of\ Dimensions} = \frac{context + input + process + product}{Number\ of\ Dimensions\ (4)}$

Explanation:

- 3.25 – 4.00 : Highly Effective
- 2.50 – 3.24 : Effective
- 1.75 – 2.49 : Moderately Effective/Adequately Effective
- 1.00 – 1.74 : Ineffective

The analysis was conducted by grouping results according to the four CIPP aspects to identify strengths and areas for improvement.

f. Reporting Results

The preparation of the evaluation report includes key findings, such as instrument feasibility, program effectiveness, recommendations for improvement, and implications of the evaluation results for further program development.

The use should be accompanied by references, and relevant modifications should be explained. Procedure and data analysis techniques should be emphasized in the literature review article. The research stages should be clearly stated.

3. Result and Discussion

The evaluation results show that the chemistry learning media lecture program, which focused on interest, is considered very effective, as it previously used an instrument with valid criteria and was reviewed by experts (0.91, very valid). The requirements are based on the average score of applying the 'Context, Input, Process, and Product (CIPP) evaluation model. Each has an actual score, such as "context" obtained by 3.75, "input" = 3.5, "process" = 3, and "product" = 3.5. The total average score is 3.43, with the criterion "very effective / very effective" achieving the highest value in the context aspect. Specifically, several indicators in each aspect receive an average score, while others do not reach the maximum.

The results also reveal several gaps, particularly in the input, process, and product phases. These results serve as a fundamental basis for teachers, study program leaders, and curriculum developers to improve the quality of lecture programs. Specifically, the findings indicate that the program has begun to support advanced learning through the development of critical thinking and problem-solving skills, while also reflecting elements of innovation learning in its project-based and interest-driven approach. However, the sustainability of these outcomes in fostering lifelong learning still requires further strengthening, particularly through improved program flexibility and continuous learning support mechanisms.

An effective program meets all four dimensions of the CIPP evaluation model (van Dulmen et al., 2023). The results illustrate the suitability of implementing CIPP, and the program meets the requirements for being called effective. The results also show that each aspect of CIPP is interconnected. This study shows that low scores in the previous aspect affect the assessment of the next element. Thus, evaluation with the CIPP model is appropriate for assessing program effectiveness because each component is interconnected, particularly the arrangement of indicators in the instrument (van Nooijen et al., 2024).

a. RQ-1: How is the design and development of CIPP-based program evaluation instruments?

Before explaining the evaluation results, the table below presents the results of instrument validation using the logical-theoretical-empirical approach (Friedenberg, 1995; Worthington & Whittaker, 2006). This approach produces valid instruments with high validity criteria, expressed as numbers and the components that follow them. The table below shows that the instrument is declared valid, with an average score of 0.95, in the very valid/high validity category (low validity criteria: $x < 0.4$; medium: $0.8 > x > 0.4$; high: $x > 0.8$). This validity also ensures that the instrument used is suitable for use as an evaluation tool. Thus, it is concluded that the CIPP model-based evaluation instrument is ideal for use as an evaluation tool.

Table 1. Data from the Analysis of Instrument Validation by Experts

Item No*	V1	V2	V3	S1	S2	S3	lo	c	n	c-1	n(c-1)	∑ S	V	Notes**
1	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
2	3	4	3	2	3	2	1	4	3	3	9	7	0.78	MV
3	4	4	3	3	3	2	1	4	3	3	9	8	0.89	HV
4	3	4	4	2	3	3	1	4	3	3	9	8	0.89	HV
5	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
6	4	3	4	3	2	3	1	4	3	3	9	8	0.89	HV
7	3	3	3	2	2	2	1	4	3	3	9	6	0.67	MV
8	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
9	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
10	3	4	3	2	3	2	1	4	3	3	9	7	0.78	MV
11	4	3	4	3	2	3	1	4	3	3	9	8	0.89	HV
12	4	3	3	3	2	2	1	4	3	3	9	7	0.78	MV
13	4	4	3	3	3	2	1	4	3	3	9	8	0.89	HV
14	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
15	4	3	3	3	2	2	1	4	3	3	9	7	0.78	MV
16	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
17	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
18	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
19	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
20	4	4	4	3	3	3	1	4	3	3	9	9	1.00	HV
Total average score													**0.91**	**HV**

*Assessment items on the validation sheet
**High Validity (HV), Moderate Validity (MV)

b. RQ-2: How effective is the program, by looking at the evaluation results of each aspect of the CIPP model?

1) The Evaluation of the "Context" of the Program

Table 2 presents the program evaluation results for the context aspect, with scale data 1-4 and 4 assessment indicators. All context aspect assessment indicators had a scale value of 4, except indicator 4, which had a scale value of 3. The total average score was 3.75, indicating it was very effective. The evaluation results suggest that the implemented program met the effectiveness requirements. Although some values are not optimal, this aspect is considered sufficient for the program's effectiveness.

Table 2. Results of Program Evaluation on the CONTEXT Aspect

Evaluation Questions "Context"	Rating Scale			
	4	3	2	1
Is the program orientation relevant to the learning outcomes of graduates of the chemistry learning media course?	√			
Does the programmer conduct a needs analysis regarding the aspect of interest before implementing the program?	√			
Is integrating programs in courses representative of students' challenges and issues in the context of digital culture?	√			
What are the personal views of students and lecturers regarding the relevance of the program to the needs of the world of work?		√		
Total	**14**			
Average	**3.75**			

Regarding the "context" aspect, the evaluator gave an average score of 3.75. This score reflects that the program is very relevant to the needs of interest-based learning and the challenges of digital culture. This relevance reflects the needs that the program meets for students. Digital culture is a pressing need for the current generation, so this program supports students' skills, primarily through projects during the program (De Witte, 2022; Trilling & Fadel, 2009). This research is relevant to one of the indicators in the "context" aspect: Is the program orientation pertinent to the learning outcomes of graduates of the chemistry learning media course? The assessment results indicate that the program aligns with the course's learning objectives. The next indicator is "What are the personal views of students and lecturers regarding the relevance of the program to the needs of the world of work?" This study also shows that the students' experiences in the program align with the current working world, equipping them with technological and critical thinking skills for future employment.

2) The Evaluation of the "Input" of the Program

Table 3 presents the evaluation data for the "input" aspect, given as Likert-scale ratings. This aspect highlights the extent to which the program runs effectively and addresses the issues of concern integrated into the selected study program. This aspect includes four assessment indicators, with indicator 4 receiving the lowest score of 2, while the others receive a maximum score of 4. With an average score of 3.5, this aspect is declared to meet the requirements and indicates that the implemented program is very effective.

Table 3. Results of the Program Evaluation on the INPUT Aspect

Evaluation Questions "Input"	Rating Scale			
	4	3	2	1
Does the program implementation use appropriate strategies, and is it relevant to the demands of graduate learning outcomes?	√			
Does implementing program strategies support students' skill and knowledge development?	√			
To what extent is the program implementation consistent and stable in achieving graduate learning outcomes?	√			
Does the program offer flexibility in duration, methods, literacy access, and mentoring to help students work on projects?			√	
Total	**14**			
Average	**3.5**			

The "input" aspect received an average score of 3.5 from the evaluator. This result shows that the implemented strategies support the development of students' skills. In the assessment, one indicator received a low score of 2: "Does the program have flexibility regarding time duration, methods, literacy access, and mentoring that helps students work on projects?" This indicator highlights the program's flexibility, which still needs improvement, especially regarding the implementation time. Through document analysis, the evaluator found that students had to work on either one project per week or on projects every week. Working on short-term projects is considered ineffective for students, and this can lead to cognitive overload. This situation can hinder student performance, affecting mental models, perceptions, memory, and students' interests (Asmussen et al., 2023). Although flexibility in duration and guidance needs improvement, these findings remain highly relevant to program flexibility, an indicator of program effectiveness (Barke et al., 2012).

3) The Evaluation of the "Process" of the program

Table 4 is the evaluation result for the "process" aspect, which consists of 4 assessment indicators. This aspect focuses on whether the program activities are running according to plan. The results show that indicators 1, 3, and 4 did not achieve the maximum score, while only indicator 2 did. Although only one indicator received the maximum score, the program had an overall average score of 3. Therefore, the program is considered adequate with respect to the process aspect because it meets the criteria.

Table 4. Results of the Program Evaluation on the PROCESS Aspect

Evaluation Questions "Process"	Rating Scale			
	4	3	2	1
Are program activities running according to plan, including the learning plan dimensions, timing, methods, and other supporting components?			√	
How is student participation in the program activities implemented?	√			
Were there any obstacles during the implementation of the program?		√		
Is there any monitoring and evaluation related to program implementation?		√		
Total	**12**			
Average	**3**			

The "process" aspect (average score of 3) highlights the obstacles to program implementation, despite monitoring and evaluation having been carried out. This aspect also received a score of 2 from the evaluator on the indicator, "Are program activities running according to plan, including the dimensions of the learning plan, time, methods, and other supporting components?” In this indicator, there is a “time” context, in which the influence of the score of 2 is determined by the previous score (input aspect). This result shows that the time element is indeed a finding in program implementation. In addition, an indicator received a score of 3, namely, "Is there any monitoring and evaluation related to program implementation?" Although this is not as bad as the score of 2 given, the suboptimal score reflects the need for improvement in the program being developed. The indicator with a score of 3 highlights monitoring of program implementation. The finding is that teachers do not closely monitor or evaluate the program, which undermines its effectiveness. Therefore, findings in the process aspect need to receive significant attention, as these results affect other assessments, especially program quality (Fitzpatrick et al., 2011).

4) The Evaluation of the “Product” of the program

Table 5 presents the results of the program evaluation for the "product" aspect, which comprises four assessment indicators. The evaluation results show that two indicators get a maximum score of 4, namely indicators 1 and 4, while others get a score of 3 (indicator 1) and 2 for indicator 3. However, the product meets the requirements, and the program is considered adequate, with a total average score of 3.5.

Table 5. Results of Program Evaluation on the PRODUCT Aspect

Evaluation Questions “Product”	Rating Scale			
	4	3	2	1
Were the program objectives achieved after the program implementation was completed?		√		
How are the students' learning outcomes after completing the program?	√			
Is the program able to improve student learning outcomes?			√	
How does the program impact students' confidence and interest in chemistry through the projects they work on?	√			
Total	**13**			
Average	**3.5**			

The "product" aspect (average score = 3.5) indicates that the program has effectively achieved its objectives, as the average score meets the requirements. However, there is an indicator that does not receive a maximum score: "Is the program able to improve student learning outcomes?" Analysis of the learning outcomes obtained shows that the 2020 batch has the lowest average score, while the highest score is obtained by students in 2022. From the graph, student learning outcomes have increased each year, with 2020 averaging around 78, 2021 averaging almost 79, and 2023 averaging around 81. On the other hand, from 2022 to 2023, there was a decline, but not by up to 2 points. Although it meets the practical requirements, this condition must be reflected in teachers' practice to maintain the program's quality and improve future learning outcomes.

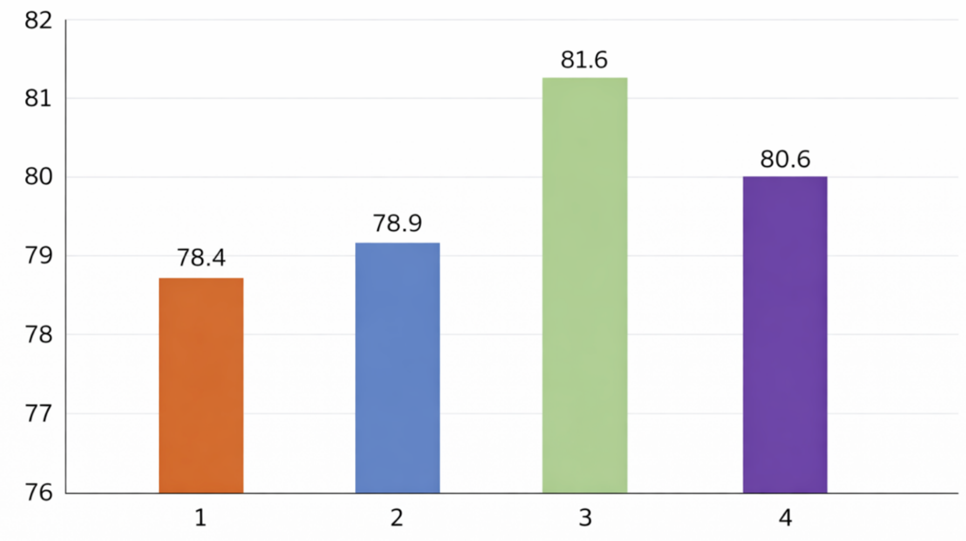

Figure 1. Description of Cognitive Learning Outcomes of Students who Participated in the Program (1 = class of 2020, 2 = 2021, 3 = 2022, 4 = 2023)

Figure 1 shows a progressive increase in students' cognitive learning outcomes from 78.4 in 2020 to a peak of 81.6 in 2022, followed by a slight decline to 80.6 in 2023. This trend indicates an overall improvement in learning effectiveness, although fluctuations suggest inconsistencies in program implementation across cohorts.

Several gaps and obstacles still need attention, as indicated by the evaluation results across the context-to-product aspects. Although the overall chemistry learning media lecture program focuses on interest, it still needs to improve to achieve a higher quality than before. For example, the program's flexibility in input, including time duration, learning methods, and access to guidance, still needs improvement. In addition, more intensive monitoring is required to overcome various obstacles in the process, as these issues affect the program's sustainability (Haynes et al., 1995). These obstacles reflect the need for more intensive monitoring of program implementation. Additional strategies are needed to improve the product aspect. Therefore, this gap provides a strong foundation for developing a program more responsive to student needs and operational challenges.

c. RQ-3: How relevant is the evaluation result to the CIPP Model structure?

The Context, Input, Process, and Product (CIPP) evaluation model was employed in this study as a comprehensive framework to assess the effectiveness of an interest-focused lecture program in chemistry learning within higher education. The application of the CIPP Model in this study not only aims to measure program effectiveness but also to examine the structural relevance between empirical findings and the theoretical constructs underlying the model. Unlike prior studies that often treat the four components of CIPP as independent evaluation dimensions, this study positions them as an interconnected system that operates dynamically across stages of program implementation.

To provide a clearer analytical structure, Table 6 presents the alignment between empirical evaluation results and the functional roles of each CIPP component.

Table 6. Alignment Between Evaluation Results and CIPP Structural Components

CIPP Component	Empirical Score	Structural Role	Observed Issue	Inter-Component Effect
Context	3.75	Needs identification & program relevance	Minor gap in job-market alignment	Influences Input design
Input	3.50	Strategy, resources, flexibility	Low flexibility in duration & mentoring	Affects Process efficiency
Process	3.00	Implementation fidelity	Time constraints, weak monitoring	Impacts Product outcomes
Product	3.50	Outcome achievement	Variation in learning outcomes	Reflects cumulative impact

As shown in Table 6, the evaluation results empirically validate the structural logic of the CIPP model. The context component, which serves as the foundational layer, achieved the highest score (3.75), indicating strong alignment between the program design and the needs of students in the digital era. This suggests that the initial stage of needs identification and program orientation has been effectively executed (Trilling & Fadel, 2009). However, the presence of a minor gap in alignment with real-world job demands indicates that contextual analysis must extend beyond academic relevance to include professional readiness and labor market expectations.

The input component, with a score of 3.50, reflects that the program has implemented appropriate strategies and resources to support learning. Nevertheless, the identified limitation in program flexibility particularly regarding project duration and mentoring access reveals a critical structural weakness. Within the CIPP framework, input serves as a bridge between planning and execution; therefore, any inefficiency at this stage has direct implications for subsequent components (Ávila & Fernández, 2025). This finding supports the argument that the CIPP model is capable of identifying not only performance levels but also structural vulnerabilities within program design.

The process component recorded the lowest score (3.00), highlighting challenges in implementation fidelity. These challenges include time constraints, inconsistencies in execution, and insufficient monitoring mechanisms. Importantly, these issues are not isolated but are directly influenced by limitations in the input stage (Ávila & Fernández, 2025). This demonstrates a causal relationship within the CIPP structure, where inadequate planning or resource allocation leads to suboptimal implementation. Such findings reinforce the model's strength in capturing dynamic interactions among evaluation components rather than treating them as static variables.

The product component, with a score of 3.50, indicates that the program has generally achieved its intended outcomes. However, the observed variability in student learning outcomes suggests that effectiveness is not uniformly distributed. This variability can be interpreted as the cumulative effect of weaknesses identified in earlier stages, particularly in input and process. Therefore, the product dimension in the CIPP model should not be viewed merely as an endpoint, but as a reflection of the entire evaluative process.

To further examine these interrelationships, Table 7 presents the structural dependency matrix within the CIPP framework.

Table 7. Structural Dependency Matrix in CIPP Evaluation

From Component	To Component	Type of Influence	Evidence from Findings	Impact Level
Context → Input	Design alignment	Needs-driven strategy formulation	Strong alignment with curriculum	High
Input → Process	Operational execution	Limited flexibility affects implementation	Time constraints observed	High
Process → Product	Outcome realization	Inefficient execution reduces optimal results	Variation in outcomes	Medium-High
Context → Product	Indirect influence	Strong relevance supports overall success	High average scores	Medium

Table 7 provides further evidence that the CIPP model operates as a hierarchical and dynamic system. The strongest relationship is observed between input and process, where limitations in flexibility directly impact implementation efficiency. This finding is particularly significant because it highlights that improving program effectiveness requires intervention at the structural level rather than merely addressing symptoms at the outcome level.

Furthermore, the relationship between process and product confirms that the quality of implementation is a key determinant of program success. Even when the program design is strong (as indicated by the context score), poor execution can limit the achievement of optimal outcomes. This underscores the importance of continuous monitoring and adaptive management within educational programs.

The indirect relationship between context and product also reveals that strong foundational alignment contributes to overall program success, even if intermediate stages experience challenges. This suggests that a well-designed program can partially compensate for implementation weaknesses, although such compensation is unlikely to sustain long-term effectiveness without structural improvements.

In conclusion, the findings demonstrate that the evaluation results are highly relevant to the theoretical structure of the CIPP model. The model not only provides a comprehensive framework for assessing program effectiveness but also enables a deeper understanding of the causal and structural relationships among evaluation components (Kusumaningtyas et al., 2024). This confirms that the CIPP model is particularly suitable for evaluating complex, interest-based educational programs in higher education, where multiple interacting factors influence program outcomes.

d. Recommendations, Limitations, and New Perspectives

The findings of this study indicate that the CIPP model offers a robust and systematic framework for evaluating educational programs, particularly those based on student interest and engagement. Beyond its role as an evaluation tool, the model also facilitates the development of actionable recommendations by identifying specific strengths and weaknesses across program components. This capability is essential for supporting evidence-based decision-making and continuous program improvement.

To operationalize these insights, Table 8 summarizes the key recommendations derived from each component of the CIPP model.

Table 8. Recommendations Based on CIPP Evaluation Findings

Aspect	Key Issue	Recommendation	Expected Impact
Context	Limited job-market alignment	Strengthen industry linkage & curriculum mapping	Higher relevance
Input	Lack of flexibility	Redesign project duration & mentoring system	Reduced cognitive load
Process	Weak monitoring	Implement real-time formative evaluation	Improved execution quality
Product	Outcome variability	Standardize assessment & feedback mechanism	More consistent outcomes

As presented in Table 8, each recommendation is directly linked to specific issues identified during the evaluation process. In the context aspect, although the program demonstrates strong alignment with academic objectives, its connection to industry and labor market demands remains limited. Strengthening this alignment through collaboration with industry stakeholders and updating curriculum frameworks is essential to enhance program relevance and graduate employability (Astutik et al., 2024).

In the input aspect, the lack of flexibility in program design particularly in terms of project duration and mentoring access emerges as a critical issue. This rigidity can lead to increased cognitive load among students, potentially reducing learning effectiveness and engagement (Himawan et al., 2024). Therefore, redesigning the program structure to allow more flexible timelines and improved access to guidance is necessary to create a more supportive learning environment.

The process aspect highlights the need for improved monitoring and evaluation mechanisms. The current evaluation approach is largely retrospective, limiting its ability to provide timely feedback during program implementation. Integrating real-time formative evaluation can address this limitation by enabling continuous monitoring and immediate corrective actions. This shift from summative to formative evaluation represents a significant advancement in program management and aligns with contemporary educational evaluation practices.

In the product aspect, variability in student learning outcomes indicates inconsistencies in program effectiveness. Standardizing assessment methods and feedback mechanisms can help reduce this variability and ensure more equitable learning outcomes across cohorts (Prastikawati et al., 2024). Additionally, incorporating multiple assessment strategies can provide a more comprehensive evaluation of student performance.

Despite these contributions, this study has several limitations that must be acknowledged. Table 9 outlines these limitations along with suggested directions for future research.

Table 9. Limitations and Future Research Directions

Dimension	Identified Limitation	Implication	Future Research Direction
Instrument	Limited validation (expert judgment only)	Potential bias	Conduct CFA/EFA & reliability testing
Scope	Single institution & program	Low generalizability	Multi-site comparative studies
Methodology	Quantitative-dominant approach	Limited depth of insight	Mixed-methods & triangulation

Dimension	Identified Limitation	Implication	Future Research Direction
Model Application	Post-hoc evaluation focus	Lack of real-time feedback	Integrate formative CIPP model

As shown in Table 9, the primary limitation of this study lies in the validation of the evaluation instrument. The reliance on expert judgment, although valuable, may introduce subjectivity and limit the robustness of the instrument. Future research should incorporate statistical validation techniques such as exploratory and confirmatory factor analysis to enhance construct validity and reliability.

Another limitation concerns the scope of the study, which is restricted to a single program within one institution. This limitation affects the generalizability of the findings, as different educational contexts may yield different results. Therefore, future studies should adopt multi-site and cross-institutional designs to provide more comprehensive and generalizable insights.

The methodological approach used in this study is predominantly quantitative, which may not fully capture the complexity of student experiences and program dynamics. Integrating qualitative methods, such as interviews and observations, can provide deeper insights into the processes underlying evaluation results (Neyazi et al., 2016). A mixed-methods approach would allow for a more holistic understanding of program effectiveness.

Finally, the application of the CIPP model in this study is primarily post-evaluative, focusing on outcomes after program completion. While this approach is useful for assessing effectiveness, it limits the model's potential to support ongoing improvement. Future research should explore the integration of formative evaluation within the CIPP framework, enabling real-time feedback and adaptive program management.

In conclusion, this study demonstrates that the CIPP model is not only effective as an evaluation framework but also valuable as a strategic tool for program development. By linking evaluation results to actionable recommendations and identifying clear directions for future research, the model supports continuous improvement and innovation in educational practice.

4. Conclusion

The study results indicate that the Context, Input, Process, Product (CIPP) model can effectively evaluate an interest-focused lecture program in chemistry learning. Macro evaluation shows this program has high effectiveness, with scores of 3.75 (Context), 3.5 (Input), 3.0 (Process), and 3.5 (Product). Evaluation with this model enables a comprehensive analysis of program aspects, from planning and implementation to the results achieved against program objectives. Although considered successful, the review also identified several weaknesses in program implementation, particularly in the learning process and flexibility, which could affect future program sustainability. This study also shows that this model can be applied in the context of an interest-focused program, with a deeper description of how each evaluation aspect influences the others.

This study contributes to the development of the CIPP model by demonstrating its capacity not only to evaluate program effectiveness but also to capture the extent to which educational programs support advanced learning, innovation learning, and lifelong learning. The findings indicate that while the program has successfully fostered higher-order thinking and innovative learning practices, its role in sustaining lifelong learning remains an area for further development. Therefore, future program design should explicitly integrate these three learning paradigms to ensure sustainable and transformative educational outcomes.

The goal is to capture the dynamics of students' experiences in the program, not just evaluation with numerical data. The implications of this study suggest that increasing the effectiveness of interest-focused programs can be achieved by adjusting the flexibility of project duration, strengthening access in mentoring contexts, and implementing more systematic and intensive monitoring strategies. Thus, further studies can develop more diverse evaluation methods by combining the CIPP model with other assessment techniques to obtain more accurate and relevant evaluation results that meet the needs of modern learning.

5. References

Akhtar, S., Nawaah, D., & Jafar, S. H. (2024). Ethical empowerment: Examining the impact of business education in Indian universities through the lens of international standards and CIPP model. *The International Journal of Management Education*, *22*(3), 101066.

Alquraan, M. F., Alazzam, S., & Farhat, D. (2025). Assessing the structural relationships between the CIPP model components in teacher education program. *Asian Education and Development Studies*, *14*(1), 103–114.

Asmussen, G., Rodemer, M., & Bernholt, S. (2023). Blooming student difficulties in dealing with organic reaction mechanisms - an attempt at systemization. *Chemistry Education Research and Practice*, *24*(3), 1035–1054. https://doi.org/10.1039/d2rp00204c

Astutik, I., Widiati, U., Ratri, D. P., Jonathans, P. M., Nurkamilah, N., Devanti, Y. M., & Harfal, Z. (2024). Transformative Practices: Integrating Automated Writing Evaluation in Higher Education Writing Classrooms - A Systematic Review. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 423–441. https://doi.org/10.23917/ijolae.v6i3.23675

Ávila, J. P. Q., & Fernández, M. B. (2025). Evaluation of a teacher pipeline program: Choosing teacher education based on self-confidence and support. *Teaching and Teacher Education*, *153*, 104842.

Barke, H.-D., Harsch, G., & Schmid, S. (2012). Essentials of Chemical Education. In *Angewandte Chemie International Edition, 6(11), 951–952.* Springer.

Barke, H.-D., Hazari, A., & Yitbarek, S. (2009). *Misconceptions in Chemistry (Addresing Perceptions in Chemical Education)*. Sense Publisher. https://doi.org/10.1007/978-3-540-70989-3_2

De Witte, M. (2022). Gen Z are not ‘coddled.’They are highly collaborative, self-reliant and pragmatic, according to new Stanford-affiliated research. *Stanford News*.

Dizon, A. G. (2023). Historical development of CIPP as a curriculum evaluation model. *History of Education*, *52*(1), 109–128.

Espinosa, A. A., Monterola, S. L. C., & Punzalan, A. E. (2013). Career-oriented performance tasks: Effects on students’ interest in Chemistry. *Asia-Pacific Forum on Science Learning and Teaching*, *14*(2).

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program Evaluation (Alternative Approaches and Practical Guidelines* (4th ed.). Pearson.

Friedenberg, L. (1995). Psychological testing: Design, analysis, and use. In *(No Title)*.

Harefa, N., Sinaga, M., & Silaban, S. (2020). Students perception and interest on chemistry: Learning evaluation integrated quizziz media. *Jurnal Pendidikan Kimia*, *12*(3), 143–150. https://doi.org/10.24114/jpkim.v12i3.21163

Haynes, S., Richard, D., & Kubiancy, E. (1995). Content validity in psychosocial assessmen: A functional approach to concept and method. *Psychological Assessment*, *7*(3), 238–247.

Himawan, R., Hermanto, H., Nurgiyantoro, B., Suyono, S., Widyartono, D., Purwanti, A., … Pogadaev, V. A. (2024). ITEMAN-Based Evaluation of End-of-Semester Assessment Items: A Case Study of Language Test in Indonesian School Context. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 387–402. https://doi.org/10.23917/ijolae.v6i3.23254

Irene, E. A. (2023). Evaluation of Teacher Education Curricula and its relevance to licensure examination using Context , Input , Process and Product (CIPP) model. *Social Sciences & Humanities Open*, *8*(1), 100607. https://doi.org/10.1016/j.ssaho.2023.100607

Kain, C., Koschmieder, C., Matischek-Jauk, M., & Bergner, S. (2024). Mapping the landscape: A scoping review of 21st century skills literature in secondary education. *Teaching and Teacher Education*, *151*(July), 104739. https://doi.org/10.1016/j.tate.2024.104739

Kusumaningtyas, D. A., Manyunu, M., Kurniasari, E., Awalin, A. N., Rahmaniati, R., & Febriyanti, A. (2024). Enhancing Learning Outcomes: A Study on the Development of Higher Order Thinking Skills based Evaluation Instruments for Work and Energy in High School Physics. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 14–31. https://doi.org/10.23917/ijolae.v6i1.23125

Kwarikunda, D., Schiefele, U., Ssenyonga, J., & Muwonge, C. M. (2020). The Relationship between Motivation for, and Interest in, Learning Physics among Lower Secondary School Students in Uganda. *African Journal of Research in Mathematics, Science and Technology Education*, *24*(3), 435–446. https://doi.org/10.1080/18117295.2020.1841961

Kwarikunda, D., Schiefele, U., Ssenyonga, J., & Muwonge, C. M. (2021). Secondary school students' motivation profiles for physics learning: Relations with cognitive learning strategies, gender, attitudes and individual interest. *African Journal of Research in Mathematics, Science and Technology Education*, *25*(2), 197–210.

Kyriazos, T. A., & Stalikas, A. (2018). Applied Psychometrics: The Steps of Scale Development and Standardization Process. *Psychology*, *09*(11), 2531–2560. https://doi.org/10.4236/psych.2018.911145

Lee, S., Shin, J., & Lee, S. (2019). *How to execute Context , Input , Process , and Product evaluation model in medical health education*. 1–8.

Lin, H. shyang, Hong, Z. R., & Chen, Y. C. (2013). Exploring the Development of College Students' Situational Interest in Learning Science. *International Journal of Science Education*, *35*(13), 2152–2173. https://doi.org/10.1080/09500693.2013.818261

Martin, I., & Carey, J. C. (2012). Evaluation capacity within state-level school counseling programs: A cross-case analysis. *Professional School Counseling*, *15*(3), 2156759X1201500302.

Meiklejohn, S., Anderson, A., Brock, T., Kumar, A., Maddock, B., Wright, C., Walker, L., & Kent, F. (2023). The utility of an interprofessional education framework and its impacts upon perceived readiness of graduates for collaborative practice. A multimethod evaluation using the context, input, process, product (CIPP) model. *Nurse Education Today*, *121*, 105707.

Neyazi, N., Arab, M., Farzianpour, F., & Majdabadi, M. M. (2016). Evaluation of selected faculties at Tehran University of Medical Sciences using CIPP model in students and graduates point of view. *Evaluation and Program Planning*, *59*, 88–93.

Ødegård, A., Strype, J., Lindqvist, S., Vasset, F., Willumsen, E., Iversen, H. P., Laake, P., & Almås, S. H. (2025). Teachers' perceptions of interprofessional collaboration in higher education: Contributions to a framework. *Social Sciences & Humanities Open*, *11*, 101300.

Prameswari, N. S., Cahyono, A., Subiyantoro, S., & Haryanto, E. (2024). Visual literacy study: influence of Instagram on interest in learning photography and its relation to gender. *Media Practice and Education*, *25*(3), 237–249.

Prastikawati, E. F., Adeoye, M. A., & Ryan, J. C. (2024). Fostering Effective Teaching Practices: Integrating Formative Assessment and Mentorship in Indonesian Preservice Teacher Education. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 230–253. https://doi.org/10.23917/ijolae.v6i2.23431

Proger, B. B. (1971). Program evaluation: The model-building game. *Journal of Learning Disabilities*, *4*(6), 292–306.

Retnawati, H. (2016). *Validitas (Reliabilitas & Karakteristik Butir)*. Parama Publishing.

Trilling, B., & Fadel, C. (2009). *21st Century Skills: Learning For Life In Our Times* (First (ed.)). Jossey-Bass.

Tuna, H., & Başdal, M. (2021). Curriculum evaluation of tourism undergraduate programs in Turkey: A CIPP model-based framework. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *29*, 100324.

van Dinther, R., de Putter, L., & Pepin, B. (2023). Features of Immersive Virtual Reality to Support Meaningful Chemistry Education. *Journal of Chemical Education*, *100*(4), 1537–1546. https://doi.org/10.1021/acs.jchemed.2c01069

van Dulmen, T. H. H., Visser, T. C., Pepin, B., & McKenney, S. (2023). Teacher and student engagement when using learning materials based on the context of cutting-edge chemistry research. *Research in Science and Technological Education*, *41*(4), 1617–1638. https://doi.org/10.1080/02635143.2022.2070147

van Nooijen, C. C. A., de Koning, B. B., Bramer, W. M., Isahakyan, A., Asoodar, M., Kok, E., van Merrienboer, J. J. G., & Paas, F. (2024). A Cognitive Load Theory Approach to Understanding Expert Scaffolding of Visual Problem-Solving Tasks: A Scoping Review. In *Educational Psychology Review* (Vol. 36, Issue 1). Springer US. https://doi.org/10.1007/s10648-024-09848-3

Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, *83*(November), 101425. https://doi.org/10.1016/j.stueduc.2024.101425

Wisudawati, A. W., Barke, H. D., Lemma, A., & Agung, S. (2022). Students' and teachers' perceptions for composition of ionic compounds. *Chemistry Teacher International*, *4*(3), 221–230. https://doi.org/10.1515/cti-2021-0032

Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

Yang, H., Yang, M., Batt, L., Xie, X., You, E., & Goff, P. (2021). A new evaluation approach for teacher preparation programs using labor market competitiveness of teacher applicants. *Teaching and Teacher Education*, *104*, 103368.

Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). Using the Context, Input, Process, and Product Evaluation Model (CIPP) as a Comprehensive Framework to Guide the Planning, Implementation, and Assessment of Service-learning Programs. *Journal of Higher Education Outreach and Engagement*, *15*(4), 57–84.

Zhao, Y., Li, W., Jiang, H., Siyiti, M., Zhao, M., You, S., Li, Y., & Yan, P. (2024). Development of a blended teaching quality evaluation scale (BTQES) for undergraduate nursing based on the Context, Input, Process and Product (CIPP) evaluation model: A cross-sectional survey. *Nurse Education in Practice*, *77*, 103976.