

Construct Validity and Reliability Analysis of a CIPP-Based Curriculum Evaluation Instrument for Outcome-Based Education (OBE)

Dhiah Fitrayati^{1*}, Heni Purwa Pamungkas¹, Renny Dwijayanti¹, Meylia Elizabeth Ranu¹, Trias Madanika Kusumaningrum¹, Khusnul Fikriyah¹
^{1*}Universitas Negeri Surabaya, Surabaya, Indonesia



DOI: <https://doi.org/10.56707/ijoerar.v4i3.187>

Sections Info

Article history:

Submitted: August 15, 2026

Final Revised: September 23, 2026

Accepted: September 24, 2026

Published: September 30, 2026

Keywords:

Curriculum

Outcome-Based Education

CIPP Model

Evaluation

Research dan Development

ABSTRACT

Objective: This study aims to develop and validate an instrument to evaluate the implementation of outcome-based education (OBE) curriculum in higher education by integrating the context, input, process, and product (CIPP) model. **Method:** The study used a research and development design that adapted the 4D model and was implemented until the development stage. The developed instrument consisted of 29 indicators distributed into four CIPP dimensions. Content validity was evaluated through expert judgment using Aiken's V, while empirical testing used partial least squares structural evaluation modeling (PLS-SEM). Evaluation of the measurement model included convergent validity, discriminant validity, and internal consistency reliability. **Results:** The four dimensions showed adequate content validity with an average Aiken's V above 0.80. Convergent validity was supported by the average variance extracted value of 0.631-0.756, while Cronbach's alpha of 0.900-0.946 and composite reliability of 0.926-0.956 indicated high internal consistency. The four dimensions have a strong and significant relationship with the evaluation construct, with the highest coefficient in the process dimension. Discriminant validity is met in most pairs of constructs, but empirical overlap is found between input and process, although all indicators have the highest cross loading on the measured construct. **Novelty:** This study produces a multidimensional instrument based on CIPP that expands the evaluation of the OBE curriculum from only measuring the achievement of learning outcomes to a diagnostic framework that integrates curriculum relevance, availability of supporting resources, implementation processes and achieved results to support the process of continuous quality improvement.

INTRODUCTION

In the higher education system, the curriculum is a strategic component that determines the effectiveness or failure of an educational program (Drakos, 2005). In the era of Society 5.0, the higher education curriculum can no longer be understood as a set of courses or academic documents, but rather as a system that continuously needs to be adjusted to developments in science, technology, the needs of the workplace, and student characteristics so that the curriculum implemented remains relevant and is able to continuously improve the quality of education. (Jackson & Bridgstock, 2021; Tight, 2024).

This paradigm shift further strengthens the implementation of the outcome-based education (OBE) paradigm in higher education. The OBE paradigm places learning outcomes as the primary orientation in educational planning, implementation, and evaluation. Within this framework, higher education needs to explicitly define the competencies students are expected to possess at the end of the program and ensure their linkage to learning objectives, learning strategies, and assessments. Thus, the OBE curriculum demands coherence between what is planned, what is implemented, and what students actually achieve. (Gurukkal, 2020; Rao, 2020; Syeed et al., 2022)

In the Indonesian context, the implementation of OBE has become an important part of higher education transformation and curriculum development. Research on OBE

implementation has been understood as an approach to curriculum reform, but its implementation still faces several challenges, particularly in formulating clear and measurable learning objectives, aligning strategies with learning outcomes, and developing an assessment system capable of measuring student competency achievement (Mufanti et al., 2024). These findings indicate that the success of OBE is not solely determined by the existence of curriculum documents based on outcomes, but also by the institution's ability to ensure implementation and achievement of outcomes through systematic monitoring and evaluation mechanisms. Other research also shows that assessment is an integral part of OBE to monitor learning outcomes and support the process of continuous improvement at the study program level (Kushari & Septiadi, 2022). Thus, the implementation of OBE requires an evaluation mechanism that can provide empirical evidence regarding the quality and effectiveness of the curriculum as a whole.

The need for comprehensive curriculum evaluation is becoming increasingly important, as the success of education is determined not only by the achievement of academic competencies, but also by their suitability to the world of work, technological developments, and global structures. Studies on graduate employability show that integrated learning experiences within the curriculum contribute to the development of graduate competencies and work readiness (Jackson & Bridgstock, 2021). Curriculum reorientation also plays a role in responding to the challenges of technological developments and the needs of the world of work (Negara et al., 2024). Therefore, OBE-based curriculum evaluation also needs to consider the suitability of the curriculum context, resource readiness, quality of implementation, and its impact on student competencies.

In the context of OBE implementation, curriculum evaluation is no longer sufficient to be carried out partially or merely to fulfill administrative requirements. Instead, it demands a holistic approach, starting with an analysis of stakeholder needs and continuing with an empirical assessment of graduates' qualifications for the workplace. Therefore, curriculum evaluation within the OBE framework requires a model that can holistically examine the relationships between curriculum components. Selecting the right curriculum evaluation model will impact the accuracy of program effectiveness measurements. The choice of curriculum evaluation model must be aligned with the objectives and characteristics of the evaluation object (Afriadi & Fitri, 2025).

For an OBE curriculum, a diagnostic gap remains when evaluation relies mainly on outcome statements, achievement scores, or administrative records: such evidence does not by itself identify which curricular conditions and teaching experiences need attention. Outcome mapping and feedback loops can show where alignment is intended, while a student-report CIPP instrument can ask how relevance, resources, delivery, and perceived outcomes are experienced in practice (Bohra et al., 2024; Derouich, 2025). These sources of evidence should be interpreted together rather than treated as interchangeable.

One model that is relevant to the characteristics of the OBE curriculum is the Context, Input, Process, Product (CIPP) model. The CIPP model was developed as an evaluation approach oriented towards providing information for decision makers and program improvement (Stufflebeam, 2003). The CIPP model has four dimensions that enable evaluators to analyze programs more comprehensively, namely the *context dimension* - identifying needs, problems, opportunities and relevance of goals; the *input dimension* -

evaluating strategies, resources, and readiness needed to achieve goals; the *process dimension* - monitoring program implementation and identifying strengths and implementation problems; the *product dimension* - assessing results, impacts, and achievement of program goals (Nouraey et al., 2020; Stufflebeam, 2003). Empirical evidence shows that the CIPP model can be used to systematically evaluate various aspects of the curriculum which include contextual suitability, input readiness, process implementation, and program outcomes (Afriadi & Fitri, 2025; Al-Shanawani, 2019; Aslan & Uygun, 2019).

The logic connecting CIPP to OBE is therefore operational. Context asks whether intended competencies match student, career, and technological needs; input asks whether staff, materials, facilities, and services can support those competencies; process asks whether teaching, assessment, participation, and feedback enact them; and product asks what competencies and benefits students perceive. The four answers can point to different improvement decisions even when a program reports the same overall outcome level (Elragal & Habibipour, 2025; Grix et al., 2005).

Although CIPP offers comprehensive evaluation coverage, its use in OBE-based curriculum evaluation presents methodological challenges. Comprehensive evaluation requires instruments capable of translating each evaluation dimension into relevant, measurable indicators that align with OBE's characteristics. Evaluation instruments must not only possess content validity based on expert assessments but also require empirical evidence that the indicators truly represent the constructs they are intended to measure. This is crucial, as curriculum evaluation data will be used as a basis for decision-making and continuous quality improvement.

Previous studies have demonstrated the use of CIPP in curriculum evaluation, but most research still focuses on evaluating program implementation using descriptive or mixed methods approaches. The CIPP model is widely used in education and other fields to assess programs and make improvements (Dizon, 2023). Among them, Aslan and Uygun (2019) used CIPP to evaluate curricula and obtain information about weaknesses in the context, input, process, and product aspects. This study demonstrates the use of CIPP for curriculum evaluation, but has not specifically developed a psychometric instrument that integrates CIPP dimensions with OBE characteristics. Thus, the CIPP model has demonstrated its ability to evaluate curricula, but the integration of CIPP with OBE principles and adequate psychometric testing to ensure construct validity and instrument reliability still require further development.

Recent CIPP applications have evaluated particular curricula or courses and used the four domains to organize findings about implementation (Rijal et al., 2026; Toosi et al., 2021; Zhang et al., 2025). The contribution tested here is more specific: whether student-report statements derived for CIPP domains in an OBE setting show acceptable content validity and measurement properties. The instrument is intended to support repeatable diagnosis, while the present analyses do not establish that it directly measures graduate competence or predicts later employment.

Based on these needs, this study aims to develop and test the validity and reliability of an OBE-based curriculum evaluation instrument using the CIPP model. Different from evaluations that only examine the achievement of learning outcomes or the suitability of curriculum documents, the instrument developed in this study is designed to evaluate the curriculum as a system measured through four dimensions: context, input, process,

and product. The context dimension is aimed at capturing the relevance of the curriculum to student needs, the world of work, and developments in science and technology; the input dimension assesses the readiness of the OBE implementation support capacity; the process dimension evaluates the implementation of learning, assessment, and feedback; while the product dimension evaluates the achievement of competencies and learning outcomes. The instrument development was then tested empirically to obtain evidence regarding construct validity and internal reliability.

RESEARCH METHOD

This research is a research and development (R&D) study aimed at developing and obtaining evidence of the validity and reliability of an OBE-based curriculum evaluation instrument using the CIPP model. The development procedure adapts the *Four -D (4D)* model developed by Thiagarajan et al (1974). Although the 4D model conceptually consists of the define, design, develop, and disseminate stages, this study is limited to the develop stage because the focus of the research is the development and testing of the psychometric quality of the instrument. Thus, the research procedure includes three stages: define, design, and develop.

Instrument development was conducted systematically, starting with the identification of construct dimensions, the development of indicators and instrument items, the assessment of content validity by experts, the revision of the instrument, and the empirical testing of the measurement model. This procedure aligns with the principles of instrument development, which place construct conceptualization, item development, content validation, and psychometric evaluation as interrelated stages in producing an instrument with adequate validity and reliability (Boateng et al., 2018).

The define stage aims to identify needs and establish a conceptual framework for an OBE-based curriculum evaluation instrument using the CIPP model. This stage is carried out through three main activities: needs analysis, literature review, and determining the evaluation focus. The literature review in the define stage was a conventional narrative review of relevant CIPP and OBE literature. The domains were specified conceptually, and the candidate items were deduced from that literature and the needs analysis. The literature review is directed at two main frameworks: the principles of OBE-based curriculum implementation and the CIPP evaluation model. The results of the literature review are then used to map dimensions and indicators relevant to curriculum evaluation. Determining the construct domain before item development is intended to ensure the indicators relate to the construct to be measured. Good instrument development must ensure that indicators and items adequately represent the construct domain to be measured (Boateng et al., 2018; Terwee et al., 2018).

The design phase aims to translate the conceptual framework obtained in the define phase into a curriculum evaluation instrument design. Activities at this stage include developing an instrument blueprint, mapping indicators for each CIPP dimension, developing statement items, determining the response scale, and developing instrument filling instructions. Item development takes into account the relationship between the indicators and the constructs being measured, as well as the clarity and readability of the statements for students as respondents. Clarity and readability of response statements are important because the validity of the instrument is also determined by the extent to which respondents understand the items according to the constructs intended by the

instrument developer (Yusoff, 2019). At this stage, each item is directed to represent students' experiences in participating in the implementation of the OBE curriculum.

The initial pool comprised 29 candidate statements distributed across context (5), input (8), process (9), and product (7). Some candidates were recorded during the development process, so the 29 statements presented in Table 2 should not be assumed to reproduce every initial wording verbatim. Deductive domain mapping and subsequent attention to how respondents understand item wording serve different purposes in scale development (Gu, 2021; Soyka & Schaper, 2024).

The development phase aims to obtain empirical evidence regarding the quality of the designed instrument. This phase is implemented through two steps: content validation through expert assessment and empirical testing of the instrument to evaluate construct validity and internal reliability. Content validation is conducted to ensure the suitability of the instrument items with the indicators and dimensions of the construct to be measured. The assessment involves two validators with expertise in curriculum evaluation, independently assessed item. They evaluated the initial draft in the first round and reassessed the revised items in the second round. The resulting Aiken's V values provide evidence of content validity based on their judgments and complement the empirical evidence obtained from student responses. Each validator assesses the draft instrument using a five-level validation sheet. The experts assess the suitability of the substance, the relevance of the indicators to the dimensions, the clarity of the statement construction, and the accuracy of the language. The content validity coefficient for each item is calculated using Aiken's V, with the formula:

$$V = \sum s / [n(c - 1)]$$

Information:

$s = r - lo$

r = score from validator

lo = lowest validity assessment number

n = number of validators

c = highest validity assessment number

Aiken's V approach in instrument development aims to quantify the level of expert agreement regarding item relevance (Ulfah et al., 2020; Yuliarto, 2021). The Aiken's V-coefficient is interpreted using the following interpretation criteria.

Table 1. Validation Data Score Interpretation Criteria

Number	Validity Criteria
0.00 – 0.40	Low
0.41 – 0.80	Medium
0.81 – 1.00	High

Source: Retnawati (2016)

After obtaining evidence of content validity, the next step was to validate the construct and reliability of the instrument using the partial least squares structural equation modeling (PLS-SEM) approach. The selection of PLS-SEM was tailored to the research objectives, which focused on developing and evaluating the quality of the measurement model for an OBE-based curriculum evaluation instrument using the CIPP

model.

The study population comprised 4,400 students at Universitas Negeri Surabaya who were undertaking their final project and had participated in the implementation of the OBE curriculum. All 4,400 students were included in the sampling frame. A sample of 354 students was selected randomly through a lottery draw.

Evaluation of the quality of the measurement model is carried out by considering the reliability of the indicators, convergent validity, discriminant validity, and internal reliability. The reliability of the indicators is evaluated based on the outer loading value of each indicator against the construct it measures, namely outer loading ≥ 0.70 (Hair et al., 2017). Indicators with outer loading values ≤ 0.70 are not automatically eliminated, but are evaluated by considering their contribution to content validity, AVE, and construct reliability (Hair et al., 2017). Convergent validity is evaluated using the average variance extracted (AVE) value with the criteria of AVE ≥ 0.5 (Hair et al., 2017). Discriminant validity is carried out using the Fornell-Larcker criterion, heterotrait-Monotrait Ratio (HTMT), and cross loadings. The internal reliability of the instrument is measured using Cronbach's alpha and composite reliability with a value of ≥ 0.7 (Hair et al., 2017).

RESULTS AND DISCUSSION

Results

Define Stage

The development of an OBE-based curriculum evaluation instrument using the CIPP model was carried out through three stages of the 4D development model, namely define, design, and develop. produced a conceptual framework for an OBE-based curriculum evaluation instrument. Based on the results of the needs analysis and literature review, curriculum evaluation using the CIPP model is constructed in four ways: context, input, process, and product (Derya & Bulent, 2016; Duan et al., 2023; Stufflebeam, 2003).

The context dimension focuses on evaluating the curriculum's suitability to student needs, career needs, and technological and information developments (Duan et al., 2023; Stufflebeam, 2003). The context dimension focuses on the curriculum's suitability to student needs, career needs, and technological and information developments.

The input dimension aims to ensure that the resources and plans prepared can support the achievement of program objectives (Derya & Bulent, 2016; Stufflebeam, 2003). The focus of the input dimension includes the availability and quality of resources supporting curriculum implementation, such as teaching staff, teaching materials, learning resources, learning facilities, technology, and academic services (Derya & Bulent, 2016; Warju, 2016).

The process dimension aims to identify needs and weaknesses in program implementation and ensure the program runs according to established plans (Duan et al., 2023; Stufflebeam, 2003). The process dimension focuses on the implementation of the learning process, student engagement, assessment, and feedback.

The product dimension aims to assess the impact and effectiveness of the program, including measuring the results achieved based on the initial objectives that have been set (Dizon, 2023; Stufflebeam, 2003). The product dimension focuses on the results of curriculum implementation, especially the achievement of students' academic and non-academic competencies.

Design Stage

During the design phase, a blueprint and draft of an OBE-based curriculum evaluation instrument were produced. The item blueprint assigns a distinct diagnostic role to each CIPP dimension. Five context items address relevance to student needs, resource readiness, and quality of learning implementation; eight input items address lecturers, materials, facilities, and support; nine process items address learning activities, participation, assessment, and feedback; and seven product items address students' perceived development and outcomes. The items were derived deductively from the literature and then reviewed by expert validators. The wording of some draft items was revised in response to their feedback. Table 2 presents the final 29 statements administered to students in this study.

Table 2. Items of the OBE-Based Curriculum Evaluation Instrument Using the CIPP Model

No	Dimensions	Statement Items
1	Context	1. The curriculum developed by the Study Program is tailored to your needs as a student (CO1) 2. The curriculum developed by the Study Program reflects the needs of the job market in your field (CO2) 3. The curriculum developed by the Study Program is relevant to the latest technological and knowledge developments in your field (CO3) 4. The curriculum developed by the Study Program helps you achieve your long-term career goals (CO4) 5. The curriculum developed by your Study Program supports the development of the competencies you need to achieve your career goals (CO5)
2	Input	6. The lecturer has sufficient competence in delivering lecture material (IN1) 7. Teaching materials (modules, power point materials, worksheets, etc.) developed by lecturers in lectures are relevant to your needs (IN2) 8. Teaching materials (modules, power point materials, worksheets, etc.) developed by lecturers in lectures are relevant to developments in science and technology (IN3) 9. The learning resources (books, journals, references, etc.) available are sufficient to support learning (IN4) 10. Laboratory facilities are adequate to support learning (IN5) 11. Library facilities are adequate to support learning (IN6) 12. The technological device facilities available are adequate to support learning (IN7) 13. Academic services such as academic guidance can support the development of your competencies to support your career goals (IN8)
3	Process	14. The learning methods used by lecturers in lectures effectively help you understand the material (PRO1) 15. The learning methods used by lecturers in lectures facilitate your active involvement (PRO2)

No	Dimensions	Statement Items
4	Product	16. The learning methods used by lecturers in lectures facilitate you to think critically (PRO3)
		17. The learning methods used by lecturers in lectures facilitate you to think creatively and innovatively (PRO4)
		18. The learning methods used by lecturers in lectures can improve collaborative (working together) competencies (PRO5)
		19. Lecturers provide feedback at the end of each learning process in class (PRO6)
		20. The lecturer provides feedback on the assignments given (PRO7)
		21. The learning process takes place according to the established schedule (PRO8)
		22. The assessment process in learning fulfills the transparency aspect (PRO9)
		23. The skills I have to achieve my career goals have improved after participating in the learning process (PROD1)
		24. The skills I have for my career goals are relevant to developments in science and technology (PROD2)
		25. My ability to think critically, analytically, and problem-solving has improved after participating in the learning process (PROD3)
		26. My ability to think creatively and innovatively has increased after participating in the learning process (PROD4)
		27. My ability to collaborate has improved after participating in the learning process (PROD5)
		28. The learning outcomes that I obtained after following the learning process were in accordance with what I expected (PROD6)
		29. In general, the implementation of lectures in the study program runs effectively (PROD7)

Develop Stage

In the development stage, the instrument draft was first evaluated by two expert validators. In addition to assessing the initial draft, the expert validators also provided suggestions for improvements to the statement wording. Based on the expert validators' suggestions, the statement items were revised accordingly. After the revisions, a re-assessment was conducted and the instrument was declared suitable for testing. The assessment results showed an average Aiken's V value of 0.850 for context, 0.813 for input, 0.819 for process, and 0.821 for product (see Table 3). Thus, all dimensions obtained an average coefficient above 0.80.

Table 3. Results of the Content Validity Analysis of the Initial Draft

No.	Dimensions	Question Items	R1	R2	Aiken's V coefficient	Category
1	Context	CO1	5	4	0.875	High
		CO2	4	4	0.750	Medium

No.	Dimensions	Question Items	R1	R2	Aiken's V coefficient	Category
2	Input	CO3	4	5	0.875	High
		CO4	5	4	0.875	High
		CO5	4	5	0.875	High
		<i>Average</i>			0.85	High
		IN1	4	4	0.750	Medium
		IN2	4	4	0.750	Medium
		IN3	4	5	0.875	High
		IN4	4	4	0.750	Medium
		IN5	4	4	0.750	Medium
		IN6	5	4	0.875	High
		IN7	4	5	0.875	High
		IN8	5	4	0.875	High
3	Process	<i>Average</i>			0.813	High
		PRO1	4	4	0.750	Medium
		PRO2	4	5	0.875	High
		PRO3	4	4	0.750	Medium
		PRO4	5	4	0.875	High
		PRO5	4	4	0.750	Medium
		PRO6	5	4	0.875	High
		PRO7	4	5	0.875	High
		PRO8	4	5	0.875	High
		PRO9	4	4	0.750	Medium
4	Product	<i>Average</i>			0.819	High
		PROD1	4	5	0.875	High
		PROD2	5	4	0.875	High
		PROD3	5	4	0.875	High
		PROD4	4	4	0.750	Medium
		PROD5	4	4	0.750	Medium
		PROD6	4	4	0.750	Medium
		PROD7	4	5	0.875	High
		<i>Average</i>			0.821	High

After obtaining evidence of content validity, the quality of the measurement model was evaluated using PLS-SEM. The evaluation included indicator reliability, convergent validity, discriminant validity, and internal reliability.

The convergent validity test results showed an AVE value of 0.714 for context, 0.631 for input, 0.671 for process, and 0.756 for product. Thus, all constructs had an AVE ≥ 0.50 , thus fulfilling convergent validity.

The results of the indicator reliability test showed outer loading values ranging between 0.691 and 0.897 (see Table 4). The lowest value was found in PRO8 (0.691), while several indicators reached the highest value of 0.897. Considering that the AVE values for all constructs exceeded 0.50 and the indicators' substantive representation of the constructs, all indicators were retained in the model. This is in accordance with the opinion of Hair et al.(2017) that indicators with outer loading values ≤ 0.70 are not automatically eliminated, but are evaluated by considering their contribution to content validity, AVE, and construct reliability.

Table 4. Convergent Validity Test Results

Dimensions	AVE	Item	Loading Factor	Information
Context	0.714	CO1	0.789	Valid
		CO2	0.829	Valid
		CO3	0.860	Valid
		CO4	0.863	Valid
		CO5	0.874	Valid
Input	0.631	IN1	0.798	Valid
		IN2	0.798	Valid
		IN3	0.859	Valid
		IN4	0.786	Valid
		IN5	0.760	Valid
		IN6	0.744	Valid
		IN7	0.749	Valid
		IN8	0.855	Valid
Process	0.671	PRO1	0.802	Valid
		PRO2	0.829	Valid
		PRO3	0.888	Valid
		PRO4	0.897	Valid
		PRO5	0.871	Valid
		PRO6	0.842	Valid
		PRO7	0.801	Valid
		PRO8	0.691	Valid
		PRO9	0.724	Valid
Product	0.756	PROD1	0.856	Valid
		PROD2	0.897	Valid
		PROD3	0.863	Valid
		PROD4	0.897	Valid
		PROD5	0.878	Valid
		PROD6	0.850	Valid
		PROD7	0.843	Valid

Discriminant validity in this study was evaluated using *the Fornell-Larcker Criterion* , HTMT, and cross-loading (Hair et al., 2017). The results of the cross-loading analysis showed that all indicators had the highest loading on the construct they automatically measure compared to other constructs. This finding supports the indicator's distinguishability at the individual level.

However, the results of *the Fornell-Larcker Criterion* test (see table 5) indicate a closeness between the input and process constructs. The correlation between input and process is $0.850 \geq$ the square root of the AVE of input of 0.795 and process of 0.891. The HTMT test value (see table 6) confirms this closeness. Using a threshold of 0.90, for conceptually close constructs, only the input-process pair is above the criterion. The HTMT results show a 95% confidence interval for input-process of 0.852-0.951. Although this interval does not include the value of 1.00, the HTMT value of 0.913 indicates that the empirical difference between input and process needs to be interpreted with caution.

Table 5. Fornell Larcker-Criterion Test Results

Dimensions	CIPP	Context	Input	Process	Product
Context	0.831	0.845			
Input	0.925	0.709	0.795		
Process	0.941	0.696	0.850	0.891	
Product	0.919	0.718	0.773	0.820	0.870

Table 6. Results of Heterotrait-Monotrait Ratio (HTMT) Test

	HTML	Significance	
		2.5%	97.5%
Context – Input	0.772	0.692	0.841
Context – Process	0.752	0.672	0.836
Context – Product	0.774	0.663	0.851
Input – Process	0.913	0.852	0.951
Input – Product	0.826	0.772	0.900
Process – Product	0.871	0.813	0.920

Crossloading analysis shows that all input indicators still have the highest loadings on the input construct, and all process indicators have the highest loadings on the process construct. Thus, no indicators were found that clearly represent other constructs more than the construct they measure. Overall, the results support discriminant validity at the indicator level and across most construct pairs. However, there is empirical overlap between the input and process dimensions, indicating the need for further refinement of the conceptual boundaries and indicator formulations of both dimensions.

Reliability testing results indicate that all constructs have high internal consistency. Cronbach's alpha and composite reliability values (see Table 7) for each construct exceeded the 0.7 criterion, thus supporting the instrument's internal consistency reliability.

Table 7. Results of Construct Reliability Testing

	Cronbach's Alpha	Composite Reliability	Information
Context	0.900	0.926	Reliable
Input	0.916	0.932	Reliable
Process	0.938	0.948	Reliable
Product	0.946	0.956	Reliable

The model estimation results show that the four dimensions have a positive and significant relationship with the OBE-based curriculum evaluation construct using the CIPP model. The coefficients obtained are context = 0.831, input = 0.925, process = 0.941 and product = 0.919 with a *p-value* < 0.001 (see table 8 and Figure 1). This indicates that each dimension contributes significantly in constructing the OBE-based curriculum evaluation instrument using the CIPP model.

Table 8. Path Coefficient Test Results

Dimensions	Path Coefficient	t- Statistics	p-values
Context	0.831	36.970	0.000
Input	0.925	63.247	0.000
Process	0.941	107.475	0.000
Output	0.919	55.950	0.000

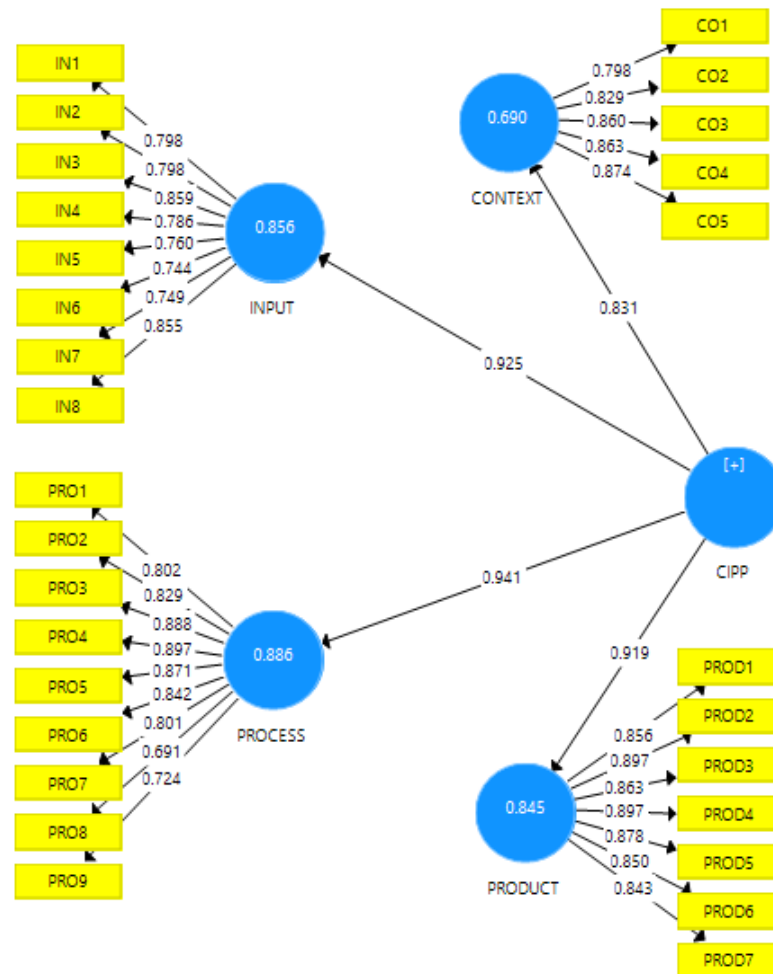


Figure 1. Construction Instrument Evaluation Curriculum OBE based with Use CIPP Approach

Discussion

The main findings of this study resulted in an OBE-based curriculum evaluation instrument constructed based on the four dimensions of the CIPP model. The development results indicate that the four CIPP dimensions can be operationalized into 29 indicators to evaluate the curriculum from a student perspective. These findings support the use of the CIPP model as an evaluation framework for evaluating the implementation of a study program's curriculum, including relevance to needs (context), adequacy of resources (input), quality of implementation (process), and results obtained (product).

This finding is consistent with the research of Duan et al.(2023) which integrates CIPP and OBE in developing a higher education evaluation indicator system in the field of engineering. This study shows that the CIPP framework can be used to build an evaluation system that connects context, input, process, and outcomes in the

implementation of OBE. Other research findings also show that the four components of CIPP can be used to systematically evaluate the fulfillment of OBE principles (Rosita et al., 2025). In addition, the CIPP model can also be used to identify the alignment of the curriculum with needs, supporting resources, learning implementation, and program outcomes (Dau & Kasiati, 2026).

The alignment of the CIPP model with OBE plays a crucial role in evaluating higher education curricula. OBE curricula are outcomes-oriented, and the use of the CIPP model can provide sufficient information on why these outcomes have or have not been achieved. This is because student outcomes are influenced by various factors. These factors can originate from the context, input, process, and product domains of curriculum implementation.

Content Validity of the Instrument

The expert judgment results provide initial evidence regarding the instrument's content validity. The average Aiken's V across the four dimensions was above 0.80. This Aiken's V value represents the final score obtained after the draft underwent a revision and revalidation process based on input from expert judgment. This process is crucial to maintaining the quality of the instrument's items in representing the conceptual domain to be measured.

Content validity findings also indicate that the four CIPP dimensions can be translated into indicators relevant to students' experiences in implementing the OBE curriculum. The context dimension measures curriculum relevance, while the input dimension maps resources and implementation readiness. The process dimension measures the learning implementation experience, while the product dimension maps the outcomes perceived by students. This separation of dimensions allows the instrument to be used to identify areas of the curriculum that still need improvement.

For interpretation and improvement, the four scores answer different questions: a weak context rating points to the relevance of intended outcomes; a weak input rating points to support capacity; a weak process rating points to what students encounter in teaching and assessment; and a weak product rating points to perceived attainment. These are prompts for further inquiry rather than automatic evidence of the cause of a problem. In particular, student reports of product should be checked against direct learning-outcome evidence where it is available (Bohra et al., 2024; Gu, 2021).

Construct Validity of Measurement Model

Empirical testing shows that all constructs meet convergent validity criteria based on AVE values. The test results show an AVE value for context of 0.714, for input 0.631, for process 0.671, and 0.756 for product. Thus, all AVE values exceed the 0.50 criterion, indicating that each construct is able to explain more than half of the variance of the indicators on average.

The outer loading value also indicates that all indicators have an adequate relationship with the construct they measure. The outer loading value ranges from 0.691 to 0.897. The lowest value was found in the process construct, namely PRO8 (0.691), which is below the criterion of 0.70. However, based on the AVE value and composite reliability of the process construct, it has met the criteria, so the statement PRO8 is maintained. This decision is in accordance with the principle of evaluating measurement models which states that indicators with an outer loading of less than 0.7 can be maintained as long as they meet the AVE criteria (Hair et al., 2020).

Internal Consistency Reliability

The test results showed high internal consistency across all dimensions. This is demonstrated by Cronbach's alpha values ranging from 0.900 to 0.946, while composite reliability values ranged from 0.926 to 0.956. Both values provide evidence that the indicators within each dimension consistently represent the same construct.

The high internal consistency results strengthen the convergent validity results. In the context of an assessment instrument, students who rate one aspect within a dimension tend to provide consistent response patterns to other indicators within the same dimension. Therefore, the developed instrument has the potential to provide diagnostic information about the implementation of the OBE curriculum.

Discriminant Validity

The results of the discriminant validity test indicate an overlap between input and process. Based on the Fornell-Larcker criterion, the input-process correlation of 0.850 is higher than the square root of the AVE of input (0.795) and process (0.819). Furthermore, this finding is reinforced by the HTMT of 0.913, which slightly exceeds the 0.90 limit. Meanwhile, other construct pairs have HTMT values below 0.90, namely 0.772 for context-input, 0.752 for context-process, 0.774 for context-product, 0.826 for input-product, and 0.871 for process-product.

These results indicate that discriminant validity is supported for most construct pairs, but the empirical distinction between input and process is not yet fully established. This finding suggests an empirical overlap between students' perceptions of curriculum support resources and their experiences with the implementation of the learning process.

Conceptually, the closeness of input and process can be understood in the context of OBE implementation. Resources that are part of the input—such as lecturer competence, teaching materials, learning facilities, technology, and academic services—are the immediate conditions that enable the process to take place. From a student perspective, for example, lecturer competence as a resource is not always easily separated from the quality of the lecturer when implementing the learning program. Another example is the availability of learning technology can be perceived through how the technology is actually used during the learning process.

The nature of the relationship between input and process, as described previously, is also consistent with the systemic logic of CIPP, namely as an evaluation system whose components are interconnected (Duan et al., 2023; Stufflebeam, 2003). The dimensions in CIPP can also be used together to explain the quality of educational provision (Aslan & Uygun, 2019; Duan et al., 2023). In the context of CIPP, input does not stand independently of process; resources and strategies are used to organize learning, while the process shows how these resources are implemented.

However, the conceptual closeness between input and process does not mean that the input and process dimensions can be combined into a single construct. This is because the results of the cross-loading analysis indicate that all indicators still have the highest loading on the construct they measure. In other words, no input indicators were found to have a higher outer loading value than the process, and conversely, no process indicators had a higher outer loading value than the input. This suggests that the discrimination problem is more likely to lie in the empirical closeness of the two constructs as a whole rather than in the misplacement of indicators. In the context of curriculum evaluation from a student perspective, resources (input) and implementation

(process) can be experienced as interrelated phenomena in the implementation of educational programs.

The Role of Context, Input, Process, and Product in OBE Curriculum Evaluation

The model results show a strong relationship between the four dimensions and the OBE-based curriculum evaluation construct. The coefficient values obtained are 0.831 for context, 0.925 for input, 0.941 for process, and 0.919 for product, and all relationships are significant. These results indicate that the four dimensions are important components in representing the curriculum evaluation construct.

The process dimension had the highest loading on the higher-order CIPP construct, indicating a strong relationship between students' reports of curriculum implementation and the overall evaluation construct. Its items address teaching activities, student participation, assessment, and feedback—experiences through which intended OBE learning outcomes are put into practice. Research on programmatic and competency-based assessment likewise emphasizes the design and enactment of assessment opportunities (Baartman et al., 2022; Vargas et al., 2025). This loading describes a measurement relationship; it does not establish that process has a greater causal effect on curriculum outcomes than context, input, or product.

The input dimension also showed a very strong relationship. This confirms that the implementation of the OBE curriculum requires adequate resource support. Well-designed and measurable learning outcomes will not automatically result in a quality learning experience without the support of competent lecturers, learning resources, learning facilities, and academic service technology and information systems.

Meanwhile, the product dimension is highly relevant to the characteristics of the OBE curriculum, which is explicitly oriented toward predetermined outcomes. These outcomes must not only be achieved but also demonstrably demonstrated by students. These outcomes will be optimally achieved when there is a correlation between needs and the availability of resources and the implementation of quality learning.

The context dimension has the lowest coefficient compared to the other dimensions, but still shows a strong relationship. The context dimension in the OBE curriculum determines whether the developed curriculum is relevant to student needs, scientific developments, and the world of work.

Ultimately, these findings suggest that integrating CIPP into OBE curriculum evaluation results in a multidimensional evaluation system in higher education. These findings reinforce Duan's findings (2023), which assert that CIPP and OBE have complementary characteristics.

CONCLUSION

Fundamental Finding: This study produced an outcome-based education (OBE) curriculum evaluation instrument that integrates the four dimensions of the CIPP model into 29 indicators. The main findings indicate that the instrument has strong evidence for this validity, convergent validity, and internal consistency reliability. However, discriminant validity revealed empirical overlap between the input and process dimensions, although cross-loading results indicated that all indicators still had the highest loadings on the constructs they measured. Thus, the developed curriculum evaluation instrument provides a reliable multidimensional structure for evaluating OBE curricula, although the empirical boundaries between resources and learning implementation still require refinement. **Implication:** Theoretically, the research findings

strengthen the use of CIPP as a framework for evaluating OBE curricula. In the context of curriculum evaluation, OBE curriculum evaluation not only measures the achievement of learning outcomes but also explains the relevance of the curriculum, the adequacy of resources, and the quality of learning implementation. Practically, the developed instrument has the potential to be used as a diagnostic tool for study programs to identify areas that need to be maintained and developed better to support the continuous quality process. Improvement. **Limitation:** This study has several limitations. First, content validity was conducted with a limited number of expert judges (2 people). Second, discriminant validity has not yet received strong support for the input-process pair. Third, the research is limited to the development stage, so the effectiveness of the instrument as a curriculum evaluation instrument has not been tested through broader implementation. **Future Research:** Therefore, further research needs to refine the conceptual boundaries and wording of input and process indicators, involving broader expert judgement and testing in various study programs and institutions.

REFERENCES

- Afriadi, B., & Fitri, F. (2025). Curriculum evaluation model: Paradigm and implications in improving the quality of education. *Jurnal Evaluasi Pendidikan*, 16(1), 50–62.
- Al-Shanawani, H. M. (2019). Evaluation of self-learning curriculum for kindergarten using Stufflebeam's CIPP model. *Sage Open*, 9(1), 2158244018822380.
- Aslan, M., & Uygun, N. (2019). *Evaluation of preschool curriculum by stufflebeam's context, input, process and product (CIPP) evaluation model*. *Egitim ve Bilim*, 44 (200), 229–251.
- Baartman, L., van Schilt-Mol, T., & van der Vleuten, C. (2022). Programmatic assessment design choices in nine programs in higher education. *Frontiers in Education*, 7, 931980.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149.
- Bohra, N. S., Johri, A., & Wasiq, M. (2024). Systematic approach of measuring program outcomes of management postgraduate program. *Frontiers in Education*, 9, 1404946.
- Dau, M. P., & Kasiati, R. I. (2026). Evaluation of the Implementation of the Christian Religious Education Curriculum Using the CIPP Model: Alignment with the Indonesian Qualifications Framework and Outcome-Based Education. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran*, 12(1), 238–250.
- Derouich, M. (2025). Ensuring outcome-based curriculum coherence through systematic CLO-PLO alignment and feedback loops. *Discover Education*, 4(1), 486.
- Derya, K., & Bulent, A. (2016). Application of context input process and product model in curriculum evaluation: Case study of a call centre. *Educational Research and Reviews*, 11(17), 1659–1669. <https://doi.org/10.5897/err2016.2911>
- Dizon, A. G. (2023). Historical development of CIPP as a curriculum evaluation model. *History of Education*, 52(1), 109–128. <https://doi.org/10.1080/0046760X.2022.2098390>
- Drakos, J. (2005). Ideas for developing a personal EFL teaching curriculum. *The Internet TESL Journal*, XI(3), 1–3.
- Duan, P., Xiang, J., Niu, H., & Han, C. (2023). Construction of evaluation index for Chinese engineering undergraduates based on Cipp model. *Sage Open*, 13(1), 21582440221149416.
- Elragal, A., & Habibipour, A. (2025). Operationalizing education 4.0: a structured method for course materials and assessment alignment. *Frontiers in Education*, 10, 1694545.
- Grix, M., Todd, S., & Lawson, R. (2005). Towards a Measure of Individual Consumer Morality. In K. M. E. and H. Brembeck (Ed.), *E - European Advances in Consumer Research* (Vol. 7, pp. 616–622). Association for Consumer Research.

- Gu, P. Y. (2021). An argument-based framework for validating formative assessment in the classroom. *Frontiers in Education*, 6, 605999.
- Gurukkal, R. (2020). Outcome-based education: an open framework. In *Higher Education for the Future* (Vol. 7, Number 1, pp. 1–4). SAGE Publications Sage India: New Delhi, India.
- Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research*, 109, 101–110.
- Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2017). *A Primer on Partial Least Square Structural Equation Modelling (PLS-SEM)* (Second). Sage Publication.
- Jackson, D., & Bridgstock, R. (2021). What actually works to enhance graduate employability? The relative value of curricular, co-curricular, and extra-curricular learning and paid work. *Higher Education*, 81(4), 723–739.
- Kushari, B., & Septiadi, L. (2022). A learning outcome assessment information system to facilitate Outcome-Based Education (OBE) implementation. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 28(2), 238–250.
- Mufanti, R., Carter, D., & England, N. (2024). Outcomes-based education in Indonesian higher education: Reporting on the understanding, challenges, and support available to teachers. *Social Sciences & Humanities Open*, 9, 100873.
- Negara, G. A. J., Pitriani, N. R. V., & Fitriani, L. P. W. (2024). Kurikulum berbasis OBE (Outcome Based Education) dengan nilai-nilai karakter untuk meningkatkan kualitas mutu pendidikan perguruan tinggi. *Jurnal Penelitian Dan Pengembangan Pendidikan*, 8(1), 41–48.
- Nouraey, P., Al-Badi, A., Riasati, M. J., & Maata, R. L. (2020). Educational Program and Curriculum Evaluation Models: A Mini Systematic Review of the Recent Trends. *Universal Journal of Educational Research*, 8(9), 4048–4055. <https://doi.org/10.13189/ujer.2020.080930>
- Rao, N. J. (2020). Outcome-based education: An outline. *Higher Education for the Future*, 7(1), 5–21.
- Retnawati, H. (2016). *Reliabilitas, Validitas dan Karakteristik Butir Soal*. Parama.
- Rijal, A. H., Rajbanshi, R., Ghimire, P., Humagain, M., Dahal, N., Hasan, M. K., & Haque, M. N. (2026). Evaluating the bachelor of dental surgery curriculum of Kathmandu university using CIPP evaluation model: a mixed-methods study. *Frontiers in Education*, 11, 1746131.
- Rosita, R., Susilana, R., Rusman, R., & Rosliana, L. (2025). Curriculum evaluation in tourism education: Assessing the fulfillment of OBE principles. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 29(2), 240–255.
- Soyka, C., & Schaper, N. (2024). Analyzing student response processes to refine and validate a competency model and competency-based assessment task types. *Frontiers in Education*, 9, 1397027.
- Stufflebeam, D. L. (2003). *The CIPP model for evaluation, in International handbook of educational evaluation*. Springer.
- Syeed, M. M. M., Shihavuddin, A. S. M., Uddin, M. F., Hasan, M., & Khan, R. H. (2022). Outcome based education (OBE): Defining the process and practice for engineering education. *Ieee Access*, 10, 119170–119192.
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, 27(5), 1159–1170.
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. University of Minnesota.
- Tight, M. (2024). The curriculum in higher education research: A review of the research literature. *Innovations in Education and Teaching International*, 61(2), 315–328. <https://doi.org/10.1080/14703297.2023.2166560>
- Toosi, M., Modarres, M., Amini, M., & Geranmayeh, M. (2021). Context, Input, Process, and Product Evaluation Model in medical education: A systematic review. *Journal of Education and Health Promotion*, 10(1), 199.
- Ulfah, A. A., Kartono, K., & Susilaningsih, E. (2020). Validity of content and reliability of inter-

- rater instruments assessing ability of problem solving. *Journal of Educational Research and Evaluation*, 9(1), 1–7.
- Vargas, H., Arredondo, E., Heradio, R., & Torre, L. de la. (2025). Standardizing course assessment in competency-based higher education: An experience report. *Frontiers in Education*, 10, 1579124.
- Warju, W. (2016). Educational program evaluation using CIPP model. *INVOTEC*, 12(1).
- Yuliarto, H. (2021). Analisis indeks aiken untuk mengukur validitas isi instrumen komitmen tugas bermain sepakbola. *Jurnal Pendidikan Jasmani Indonesia*, 17(1), 19–23.
- Yusoff, M. S. B. (2019). ABC of response process validation and face validity index calculation. *Educ Med J*, 11(10.21315).
- Zhang, S., Bai, L., & Yang, B. (2025). Mixed teaching quality evaluation of organizational behavior course based on CIPP model. *Frontiers in Education*, 10, 1538539.

***Dhiah Fitrayati (Corresponding Author)**

Department of Economics Education Faculty of Economics and Business,
Universitas Negeri Surabaya,
Address: Jalan Ketintang, Surabaya 60231, Indonesia
Email: dhiahfitrayati@unesa.ac.id

Heni Purwa Pamungkas

Department of Economics Education Faculty of Economics and Business,
Universitas Negeri Surabaya,
Address: Jalan Ketintang, Surabaya 60231, Indonesia
Email: henipamungkas@unesa.ac.id

Renny Dwijayanti

Department of Business Education Faculty of Economics and Business,
Universitas Negeri Surabaya,
Address: Jalan Ketintang, Surabaya 60231, Indonesia
Email: rennydwijayanti@unesa.ac.id

Meylia Elizabeth Ranu

Department of Business Administration Education Faculty of Economics and Business,
Universitas Negeri Surabaya,
Address: Jalan Ketintang, Surabaya 60231, Indonesia
Email: meyliaranu@unesa.ac.id

Trias Madanika Kusumaningrum

Department of Management Faculty of Economics and Business,
Universitas Negeri Surabaya,
Address: Jalan Ketintang, Surabaya 60231, Indonesia
Email: triaskusumaningrum@unesa.ac.id

Khusnul Fikriyah

Department of Islamic Economics Faculty of Economics and Business,
Universitas Negeri Surabaya
Address: Jalan Ketintang, Surabaya 60231, Indonesia
Email: khusnulfikriyah@unesa.ac.id
