LEARNING PROGRAM EVALUATION AND TEACHING QUALITY IMPROVEMENT: A QUANTITATIVE CIPP STUDY

Desembra Sohilait[1*], Clara Samarama[2], Sarni Reniwurwarin[3], Nur Yanti Habu[4], Iwan Yusuf[5]
[1,2,3,4,5] Universitas Pattimura, Indonesia
*Corresponding Author: desembrasohilait@gmail.com

ABSTRACT
This study was motivated by the absence of a comprehensive and systematic evaluation of school learning programs designed to improve teaching quality at SMP Negeri 3 Ambon. Although the school has implemented several improvement initiatives, including teacher training, academic supervision, project-based curriculum development, and the strengthening of digital learning facilities, their overall effectiveness had not been quantitatively assessed. Therefore, this study aimed to evaluate the effectiveness of these learning programs using the Context, Input, Process, and Product (CIPP) model. A descriptive quantitative design was employed involving 20 respondents, consisting of teachers and educational staff, selected through total sampling. Data were collected using a Likert-scale questionnaire developed according to the four CIPP dimensions and analyzed through descriptive statistics and Pearson correlation using SPSS version 25. The findings revealed that the overall implementation of the learning programs was rated very good, with a mean score of 4.41. Among the CIPP dimensions, product received the highest score (4.70), followed by context (4.60), process (4.40), and input (4.30). In addition, the Pearson correlation coefficient of 0.781 ($p < 0.05$) indicated a significant relationship between learning program implementation and teaching quality improvement. The study concludes that the CIPP model is effective for comprehensively evaluating school learning programs and for identifying strengths and areas for improvement. These findings provide empirical support for school leaders and policymakers in making evidence-based decisions to enhance teaching quality. Future research is recommended to involve larger samples, multiple schools, and mixed-method approaches to produce broader and deeper evaluation results.

Keywords: CIPP model; Learning program evaluation; Quantitative evaluation; School improvement; Teaching quality

INTRODUCTION

Over the past decade, teaching quality has become a central concern in educational research and school improvement discourse. The quality of teaching is widely recognized as a major determinant of students' learning experiences and academic achievement, as well as a key indicator of school effectiveness. In response, education systems have increasingly emphasized teacher professional development, instructional leadership, curriculum responsiveness, and the integration of digital resources as strategic means of enhancing teaching quality (UNESCO, 2024; OECD, 2016; OECD, 2025). Contemporary scholarship conceptualizes teaching quality as a multidimensional construct encompassing lesson planning, classroom management, student engagement, feedback, and responsiveness to learners' needs. Studies have consistently shown that improvements in teaching quality contribute significantly to better student outcomes and stronger school performance. However, the effectiveness of school improvement initiatives depends not only on program implementation, but also on the availability of systematic evaluation mechanisms capable of assessing whether such programs are relevant, adequately supported, properly executed, and effective in practice (Gore et al., 2017; Darling-Hammond et al., 2017). In the Indonesian context, this issue has become increasingly important as schools are required to respond simultaneously to curriculum reform, teacher development demands, and digital transformation. These conditions make program evaluation essential for ensuring that school-based initiatives produce measurable contributions to teaching quality and are supported by evidence-based decision-making.

Despite the growing importance of school improvement programs, it remains uncertain whether the learning programs implemented in many schools genuinely improve teaching quality. In many

cases, schools organize teacher training, academic supervision, curriculum development, and digital support initiatives as separate activities without evaluating them as a unified program. Consequently, school leaders may find it difficult to determine whether such programs are contextually appropriate, sufficiently supported, effectively implemented, and capable of producing meaningful outcomes. This challenge is also reflected in prior research. Many evaluation studies focus on isolated aspects, such as outcomes, participant satisfaction, or resource availability, without examining the interrelationship among program needs, resources, implementation processes, and results. Other studies are limited to qualitative approaches or highly specific settings, thereby reducing their relevance for broader school-based decision-making. These limitations indicate the need for a comprehensive and practical evaluation framework that can assess educational programs holistically (Stufflebeam, 2003; Aziz et al., 2018).

One of the most widely recognized frameworks for comprehensive program evaluation is the CIPP model developed by Stufflebeam. This model evaluates programs through four dimensions: context, input, process, and product. Its major strength lies in its capacity to assess not only final outcomes but also the relevance of program objectives, the adequacy of resources, the quality of implementation, and the effectiveness of results achieved. Previous studies have applied the CIPP model across diverse educational contexts. Aziz et al. (2018) employed it to evaluate school quality, while Erdogan and Mede (2021) used it in language education. Other studies have shown that the model is effective in identifying program strengths, limitations, and areas for improvement in both educational and professional settings (Lippe & Carter, 2018; Irene, 2023; Chanthalangsy et al., 2024). These findings suggest that the CIPP model remains highly relevant for educational evaluation and institutional decision-making.

Although the CIPP model has been widely used, several gaps remain evident in the literature. First, many studies have applied the model in higher education, teacher education, or subject-specific programs rather than in whole-school learning improvement programs at the junior secondary level. Second, many evaluations have not explicitly linked program assessment to the broader issue of teaching quality improvement. Third, several studies have been limited in scale or have focused on only one dimension of evaluation, thereby reducing their usefulness for integrated school-level decision-making. This gap is particularly relevant to SMP Negeri 3 Ambon. The school has implemented several initiatives intended to improve teaching quality, including teacher training, academic coaching, curriculum development, and improvements in digital facilities. However, these initiatives have not yet been evaluated through a single comprehensive quantitative framework. As a result, the school still lacks empirical evidence regarding the overall effectiveness of its learning programs.

The purpose of this study was to evaluate the implementation of learning programs at SMP Negeri 3 Ambon using the CIPP model and to examine their relationship with teaching quality improvement. The novelty of the study lies in its integration of all four CIPP dimensions within a single quantitative evaluation at the junior secondary school level, its focus on a public school context in Ambon, and its emphasis on teaching quality as an important outcome of program evaluation. This study was limited to SMP Negeri 3 Ambon and involved teachers and educational staff directly engaged in the implementation of the school's learning programs. Since the study relied on questionnaire data, the findings reflect participants' perceptions rather than direct classroom observation or long-term student achievement data. Future research is therefore recommended to involve multiple schools, mixed-method designs, and broader evaluation indicators in order to produce a more comprehensive understanding of sustainable teaching quality improvement.

METHOD

Research Design and Approach

This study employed a quantitative descriptive approach within an evaluative research design to examine the implementation of school learning programs and their contribution to teaching quality improvement at SMP Negeri 3 Ambon. The evaluation was guided by the CIPP model—Context, Input, Process, and Product—which is widely used to assess educational programs systematically and comprehensively (Stufflebeam & Coryn, 2014). This model was selected because it enables the evaluation of a program not only in terms of its outcomes, but also in relation to its contextual relevance, resource adequacy, implementation quality, and overall results. The study was non-experimental in nature, as it did not involve treatment manipulation or group comparison. Instead, it focused on describing the current condition of program implementation based on respondents' perceptions and on identifying the extent to which the implementation of learning programs was associated with improved teaching quality. The design was considered appropriate because the purpose of the study was evaluative rather than causal, and it sought to generate empirical evidence for school-based decision-making.

Participants

The population of this study consisted of all educational personnel directly involved in the implementation of learning programs at SMP Negeri 3 Ambon. This population included 18 subject teachers and 2 educational staff members, resulting in a total of 20 participants. Because the number of accessible participants was limited and manageable, the study used a saturated sampling technique or census sampling, in which all members of the population were included as research participants. This sampling strategy was considered suitable because all participants had direct experience with the planning, implementation, and evaluation of school learning programs. Their involvement allowed the study to obtain a comprehensive picture of how the programs were perceived and implemented within the school context.

Data Collection Techniques and Instruments

The data used for this study were collected by means of a structured questionnaire administered directly to all participants. The instrument was developed based on the four dimensions of the CIPP evaluation model, namely context, input, process, and product. The questionnaire consisted of 20 items, with five items assigned to each dimension. All items were measured using a five-point Likert scale, ranging from 1 = strongly disagree to 5 = strongly agree. The context dimension measured the relevance of the learning program to school needs, policy direction, and instructional priorities. The input dimension examined the adequacy of human resources, facilities, administrative support, and other resources required for program implementation. The process dimension assessed how the program was carried out, including training activities, supervision, assistance, and reflection. Finally, the product dimension measured the perceived outcomes of the program, particularly in relation to teacher professionalism, the use of learning media, and improvement in teaching quality.

Before the questionnaire was distributed, the items were aligned conceptually with the CIPP framework to ensure that each statement reflected the intended evaluation component. The questionnaire was then administered in the school setting after permission had been obtained from the school leadership.

Table 1. Structure of the Research Instrument

Dimension	Focus of Measurement	Number of Items	Scale
Context	Program relevance, school needs, policy alignmen	5	5-point Liker
Input	Human resources, facilities, administrative and institutional support	5	5-point Liker
Process	Program implementation, supervision, coaching, reflection activities	5	5-point Liker
Product	Perceived outcomes, teaching quality improvement, professional growth	5	5-point Liker
Total		20	

Data Analysis Procedures

The collected data were coded and analyzed using SPSS version 25. Data analysis was conducted in two stages. First, descriptive statistics were used to summarize respondents' perceptions of each CIPP dimension.

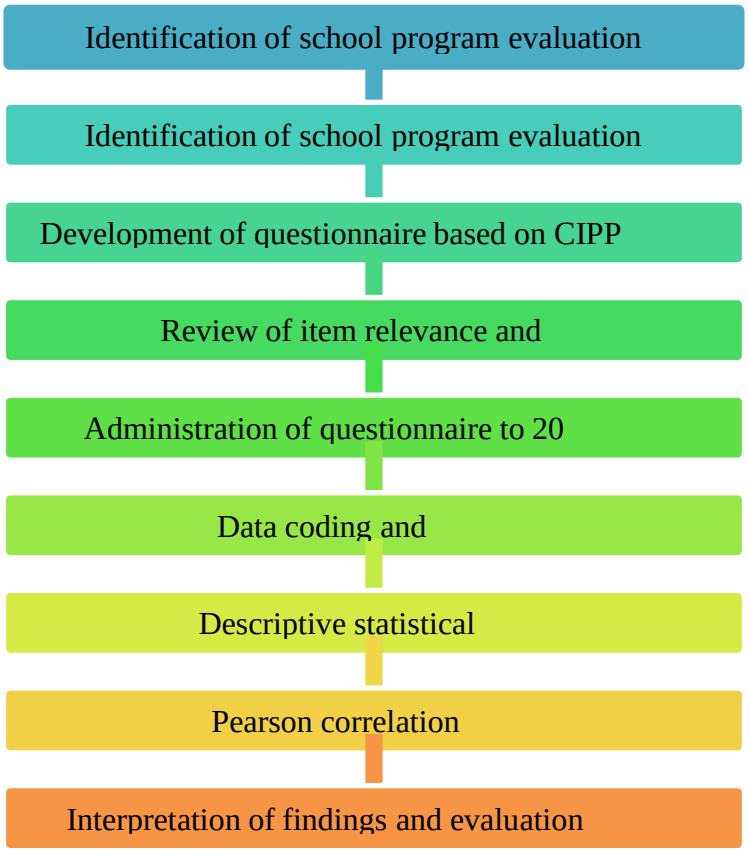


Figure 1. Research Procedure Flow

These statistics included mean scores and standard deviations, which were used to identify the general tendency and consistency of responses across the four evaluation components. Second, a Pearson correlation analysis was conducted to examine the relationship between the implementation of learning programs and teaching quality improvement. This analysis was used to determine whether better program implementation was significantly associated with better perceived teaching quality. The level of significance was set at $p < 0.05$. The overall data analysis procedure is summarized in Figure 1.

Validity, Reliability, and Ethical Considerations

Instrument validity in this study was established through content alignment with the theoretical indicators of the CIPP model and through a review of item relevance and clarity prior to data collection. This procedure was intended to ensure that the questionnaire adequately represented the four dimensions being evaluated. Reliability was examined through internal consistency analysis using SPSS. In the final manuscript, it is advisable to report the reliability coefficient, such as Cronbach's alpha, for the overall instrument and, if available, for each dimension. Ethically, this study followed the basic principles of educational research. Prior permission to conduct the study was obtained from the school. Participation was voluntary, and respondents were informed of the purpose of the study before completing the questionnaire. They were also assured that their responses would be treated confidentially and used only for academic purposes. No personal identifiers were reported in the analysis or presentation of results, thereby protecting participant anonymity and minimizing the risk of response bias.

RESULTS AND DISCUSSION

The findings of this study clearly show that the learning program implemented at SMP Negeri 3 Ambon was perceived very positively by the respondents. The descriptive results indicate an overall mean score of 4.41 with a standard deviation of 0.44, placing the program in the "excellent" category. Among the four CIPP dimensions, the Product component obtained the highest mean score (4.70), followed by Context (4.60), Process (4.40), and Input (4.30). The respondents consisted of 20 school personnel, including 15 teachers, 3 deputy principals, and 2 education staff. Most had more than 10 years of work experience (65%), while 80% held a bachelor's degree and 20% held a master's degree. These data suggest that the judgments reported in this study came from participants with substantial institutional experience and direct involvement in program implementation, which strengthens the practical credibility of the evaluation findings, even though the data remain perception-based rather than experimental.

Viewed through the CIPP framework, the pattern of findings suggests that the program was not only considered relevant to school needs, but also reasonably well supported, effectively implemented, and strongly associated with desirable outcomes. This pattern is highly consistent with the original logic of the CIPP model, which treats evaluation as a decision-oriented and improvement-oriented process rather than a mere exercise in proving program success. In that sense, the present findings do more than indicate satisfaction; they also identify where the program appears strongest and where strategic improvement is still needed. The relatively high scores across all four dimensions imply that the program was coherent across planning, resourcing, implementation, and outcomes. Viewed through the CIPP framework, the pattern of findings suggests that the program was not only considered relevant to school needs, but also reasonably well supported, effectively implemented, and strongly associated with desirable outcomes.

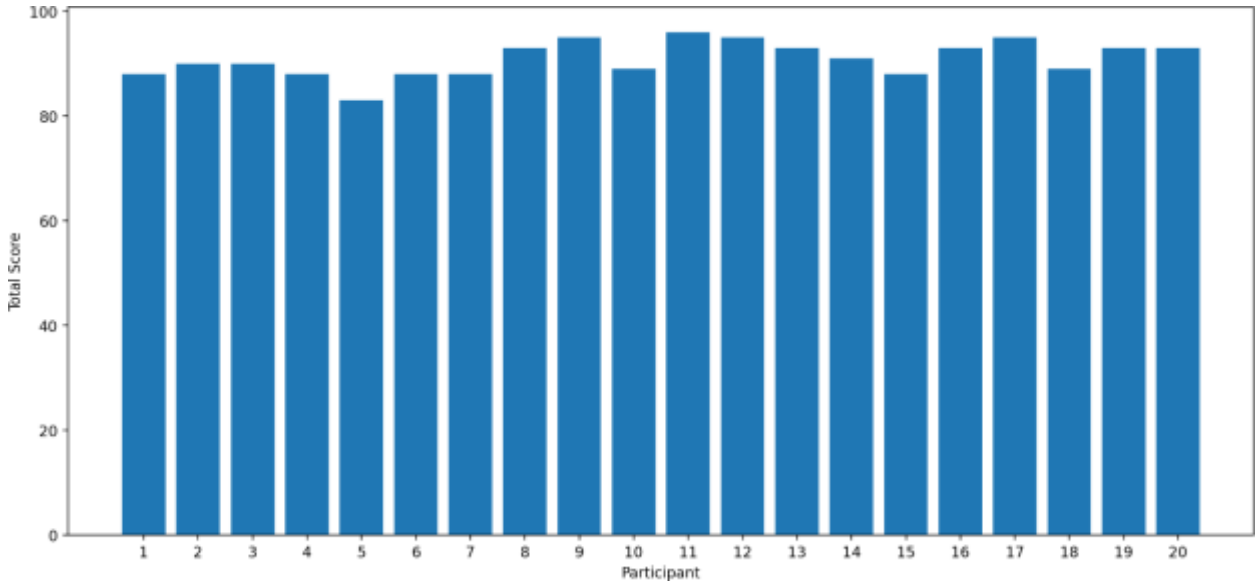


Figure 2. Participant Score Summary for CIPP Evaluation

This pattern is highly consistent with the original logic of the CIPP model, which treats evaluation as a decision-oriented and improvement-oriented process rather than a mere exercise in proving program success. In that sense, the present findings do more than indicate satisfaction; they also identify where the program appears strongest and where strategic improvement is still needed. The relatively high scores across all four dimensions imply that the program was coherent across planning, resourcing, implementation, and outcomes. At the same time, the gap between the highest and lowest dimension scores is analytically important because it reveals that the strongest perceived gains were located at the outcome level, while the most modest ratings appeared at the resource and support level. That pattern is a classic signal in program evaluation: schools may achieve strong results not because resources are perfect, but because contextual alignment, staff commitment, and implementation quality compensate for infrastructure limitations (Stufflebeam, 2003).

Table 2. descriptive statistical

Components	Number of Items	Red	Std. Deviation	Categories
Background	5	4.60	0.42	Excellent
Input	5	4.30	0.49	Good
Process	5	4.40	0.45	Good–Very Good
Product	5	4.70	0.38	Excellent
Total Score	20	4.41	0.44	Excellent

The strong overall evaluation can also be interpreted in light of broader literature on teacher professional development and school improvement. A large body of scholarship has shown that effective professional development is most likely to produce positive effects when it is content-focused, coherent with institutional goals, practical in orientation, and sustained long enough to affect classroom routines.

Garet et al. (2001) emphasized the importance of content focus, active learning, and coherence, while Desimone (2009) argued that professional development should be conceptualized not as an isolated event but as a chain linking program features, teacher learning, instructional change, and student outcomes. More recent studies reinforce that point by showing that professional development quality is associated with dimensions such as clarity, structure, collaboration, and practical relevance, and that the effects of professional development on teaching and student learning are real but uneven and highly dependent on program design. Therefore, the very favorable overall mean in this study should not be read as a generic expression of satisfaction; rather, it suggests that respondents recognized several core characteristics that international research identifies as markers of effective professional learning (Desimone, 2009; Garet et al., 2001).

The Context score of 4.60 is especially important because it indicates that the program was perceived as highly relevant to the school's real needs and to current educational priorities. In practical terms, this means respondents did not see the program as externally imposed or administratively symbolic; instead, they viewed it as aligned with the mission of improving teaching quality, strengthening teacher competence, and supporting curriculum implementation. This is a crucial foundation for program success. International evidence repeatedly shows that when professional development is closely connected to teachers' daily work and local school goals, it is more likely to be taken seriously and translated into practice. In the Indonesian context as well, recent review evidence suggests that professional development contributes most effectively to teacher quality when it is sustained, relevant, and institutionally supported. Thus, the high context score in this study can be interpreted as a sign of strong program legitimacy. The school appears to have succeeded in framing the program as a meaningful response to authentic instructional challenges rather than as a procedural obligation. That legitimacy matters because teachers are far more likely to engage deeply with a program when they believe it addresses actual classroom needs (Rahmi & Rassanjani, 2025).

The Input component, although still rated "good," obtained the lowest mean score (4.30), making it the most critical area for analytical discussion. This result suggests that human resources, school management support, and learning facilities were adequate, but not equally strong in all respects. The attached text explicitly notes several challenges, especially limited internet access and some digital devices that were not functioning optimally. This is a highly significant finding because it shows that the main constraint was not instructional willingness or program acceptance, but the infrastructure needed to support consistent implementation. This pattern corresponds closely with recent international literature on digital professional development and technology integration. Studies have shown that teacher growth in digital pedagogy depends not only on training content but also on access to functioning devices, school-level support, and a digitally ready environment. Reviews on rural and unevenly resourced school settings further demonstrate that infrastructure gaps continue to slow the translation of professional learning into stable instructional routines. Accordingly, the present study's relatively lower input score should not be dismissed as a minor technical issue. It points to a structural condition that could eventually limit sustainability if not addressed through better internet access, device maintenance, technical support, and budgeted infrastructure renewal (Montero-Mesa et al., 2023).

The Process score of 4.40 indicates that the program's implementation was perceived as effective and professionally meaningful. According to the attached summary, the process included teacher training, academic supervision, workshops, learning assistance, and reflective activities. This combination is analytically important because it reflects a job-embedded model of professional development rather than a one-shot seminar model. International meta-analytic evidence consistently shows that coaching, observation, feedback, and collaborative support have stronger effects on instructional practice than brief and disconnected forms of training. Teacher learning also tends to improve when school leadership actively supports professional growth and when collaborative structures create space for joint problem-solving. In that sense, the process findings are

one of the strongest parts of this study. They suggest that the school was not only offering training content, but also creating a process architecture that helped teachers interpret, practice, and refine what they learned. This helps explain why the product dimension later emerged as the highest-rated component. A strong implementation process often acts as the bridge between program intentions and visible professional change (Kraft et al., 2018).

The reflective dimension of the process deserves special attention. The attached text indicates that teachers experienced training and reflection as activities that improved motivation, teaching methods, and their ability to examine learning practices. This is important because reflection is not merely a supplementary activity; it is one of the mechanisms through which professional development becomes transformative. International studies describe reflective practice as a process that helps teachers interpret experience, surface assumptions, and revise classroom decisions more consciously. Structured reflection, especially when supported by prompts, collaborative dialogue, observation, or video analysis, has been shown to strengthen metacognition, alter teachers' perceptions of their own practice, and enhance self-efficacy. Therefore, the strong process score may partly reflect the school's ability to move teachers beyond compliance into reflective engagement. One explanation for the positive rating is that respondents perceived the program as intellectually and professionally usable, not just administratively required. This matters because professional development has greater impact when teachers can connect training experiences directly to concrete classroom problems and then reflect on the consequences of instructional change (Chaseley & Abercrombie, 2025).

The Product component obtained the highest mean score (4.70), and this is perhaps the clearest indicator of the program's perceived success. Respondents believed that the program had a real effect on teacher professionalism, the use of instructional media, classroom creativity, and student participation. In evaluation terms, this means the program was seen as generating outputs and early outcomes that were visible in daily practice. The prominence of the product score is consistent with evidence showing that well-designed professional learning can improve teacher knowledge, instructional quality, and, under favorable conditions, student engagement and achievement. Reviews and meta-analyses increasingly show that teacher professional development can influence student outcomes, although the size of the effect depends heavily on the quality, duration, and focus of the intervention. In the present study, the product dimension appears to capture teachers' sense that the program changed what they actually do, not merely what they know. That distinction is important. Professional development is most educationally valuable when it alters planning, instructional delivery, assessment, and interaction with learners. The high product score suggests that respondents saw such changes occurring in meaningful ways at SMP Negeri 3 Ambon (Ventista & Brown, 2023).

At the same time, the relationship among the four components reveals a more nuanced picture than a simple claim of success. The fact that Product is the highest-rated component while Input is the lowest-rated one suggests that positive outcomes may have been achieved despite imperfect material conditions. This is a noteworthy pattern. It may indicate that experienced teachers, collaborative culture, and effective school leadership were able to compensate for limitations in technology and facilities. Given that many respondents had over a decade of experience, it is plausible that their pedagogical maturity made them more capable of adapting training content even when infrastructure was not ideal. Another possible explanation is that the school's implementation routines, such as supervision and reflection, created enough internal support to reduce the negative effects of material constraints. However, an alternative interpretation should also be considered: because the data are based on participant perceptions, respondents may have been more sensitive to visible improvements in teaching practice than to persistent weaknesses in infrastructure. In other words, strong product ratings may reflect genuine change, but they may also reflect optimism shaped by commitment to the program. Both explanations are plausible, and both deserve acknowledgement

in the interpretation (Hallinger et al., 2014; Stufflebeam, 2003).

Compared with previous literature, the present findings are broadly consistent rather than deviant. First, the strong context and process scores align with literature arguing that teacher learning is most powerful when school programs are coherent, collaborative, and connected to ongoing school improvement. Second, the relatively weaker input score is consistent with many studies reporting that infrastructure and institutional conditions often lag behind instructional ambition, especially in contexts where digital access and technical maintenance remain uneven. Third, the very high product score corresponds with studies showing that teachers often report tangible gains in professionalism, instructional confidence, and classroom effectiveness when development programs are accompanied by coaching, reflection, and school-based support. Where this study differs slightly from some international findings is in the magnitude of positivity across all dimensions. Many large-scale reviews report more modest and variable effects. The uniformly high ratings in this study may therefore reflect the focused nature of a single-school program, the experience level of respondents, or the positive social climate surrounding the intervention. This does not invalidate the findings, but it suggests that they should be interpreted as strong school-specific evidence rather than as a universal benchmark (Desimone, 2009; Garet et al., 2001; Kraft et al., 2018).

The importance of these findings extends beyond the single program being evaluated. Theoretically, the results reinforce the usefulness of the CIPP model as a diagnostic framework for educational improvement. Rather than limiting evaluation to final outcomes, the model made it possible to identify a meaningful asymmetry in the program: strong relevance, good implementation, excellent outcomes, but somewhat weaker supporting resources. That diagnosis is valuable because it suggests where future intervention should be directed. In terms of contribution to the literature, the study adds school-level evidence from an Indonesian secondary school showing that positive evaluations of teaching-quality programs are closely associated with contextual fit, job-embedded implementation, and strong perceived outcomes. Practically, the results suggest three policy directions. First, school leaders should preserve the elements that produced strong process and product ratings, especially supervision, workshops, mentoring, and reflective follow-up. Second, policymakers and school managers should prioritize investment in internet access, digital devices, and maintenance systems because these are the most visible weak points in the evaluation. Third, future program cycles should integrate stronger evidence of student learning and classroom observation so that the next stage of evaluation moves beyond perception to documented instructional change (Rahmi & Rassanjani, 2025; Stufflebeam, 2003).

Several unexpected or analytically interesting aspects should also be noted. One is that the program was rated very highly even though input limitations were openly acknowledged. This may indicate resilience in the school's professional culture, where teachers continue to innovate despite infrastructure gaps. Another notable feature is the relatively low standard deviation values across components, which indicate a fairly high level of agreement among respondents. This consistency suggests that perceptions of program quality were shared rather than fragmented across staff roles. Nevertheless, a low spread can also signal a possible ceiling tendency, particularly in small-scale school surveys where respondents may be inclined toward favorable ratings. Such a pattern does not negate the findings, but it means that the evidence should be interpreted as internally coherent rather than automatically comprehensive. One explanation for the high convergence of ratings is that the program was experienced collectively and evaluated within a shared institutional culture. Another explanation is that the benefits were sufficiently visible to generate similar judgments across teachers and school leaders. These interpretations are not mutually exclusive.

The limitations of the findings must be discussed explicitly. First, the study relies on a relatively small sample from a single school, so the results cannot be generalized without caution. Second, the evidence is descriptive and perception-based; it identifies how participants judged the program, but it does not establish causal impact in the experimental sense. Third, the strongest claims concern professional practice as perceived by school staff, not independently verified changes in classroom

instruction or student achievement. International evaluation literature repeatedly warns that strong perceptions of success should ideally be complemented by observation, longitudinal follow-up, and outcome triangulation. Moreover, some professional development effects weaken over time when follow-up support is insufficient, meaning that high short-term ratings do not automatically guarantee durable long-term change. For that reason, the present findings should be considered reliable as a portrait of current staff perceptions, but provisional as evidence of lasting impact. Future studies would be stronger if they combined questionnaires with classroom observation, document analysis, student performance data, and repeated measures across semesters or academic years (Hallinger et al., 2014).

Overall, the results demonstrate that the learning program at SMP Negeri 3 Ambon was perceived as highly successful in improving the quality of teaching. The strongest evidence lies in the Product and Context dimensions, indicating that the program was both relevant and beneficial, while the Process results show that implementation was pedagogically meaningful and professionally supportive. The main area requiring strategic improvement is Input, particularly in relation to digital infrastructure and technical readiness. In conclusion, the evaluation suggests that the program has built a solid foundation for improving teaching quality, but its sustainability and future scalability will depend on whether the school and its stakeholders can strengthen the resource base that supports implementation. Thus, the study supports the view that effective school improvement is not achieved by resources alone or by training alone, but by the alignment of contextual relevance, adequate support, reflective implementation, and visible professional outcomes (Stufflebeam, 2003).

CONCLUSION

This study aimed to evaluate the implementation of learning programs at SMP Negeri 3 Ambon using the Context, Input, Process, and Product (CIPP) model and to examine their relationship with teaching quality improvement. The findings indicate that the overall implementation of the program was rated very positively, with all four CIPP dimensions falling within the good to excellent range, and with the Product and Context components emerging as the strongest aspects of the program. These results suggest that the learning program was not only relevant to the school's instructional needs, but was also effectively implemented and associated with meaningful improvements in teacher professionalism, instructional practice, and perceived teaching quality. The study contributes to educational evaluation theory by reaffirming the usefulness of the CIPP model as a comprehensive framework for diagnosing program strengths and weaknesses at the school level. In practical and policy terms, the findings provide evidence that school improvement efforts should be supported by coherent planning, ongoing supervision, reflective professional development, and stronger investment in digital infrastructure and instructional resources. Future research is recommended to involve larger samples, multiple school settings, and mixed-method or longitudinal designs in order to generate deeper and more generalizable evidence on the long-term effectiveness of learning programs in improving teaching quality.

REFERENCES

Aziz, S., Mahmood, M., & Rehman, Z. (2018). Implementation of CIPP model for quality evaluation at school level: A case study. Journal of Education and Educational Development, 5(1), 189–206. https://doi.org/10.22555/joeed.v5i1.1553

Chanthalangsy, P., Yeh, B.-I., Choi, S. J., & Park, Y. C. (2024). Program evaluation of postgraduate obstetrics and gynecology training in Lao People's Democratic Republic using the CIPP model. BMC Medical Education, 24(1), Article 44. https://doi.org/10.1186/s12909-023-04942-6

Chaseley, T. L., & Abercrombie, S. (2025). Using scaffolds to support preservice teachers' reflective practice. Frontiers in Education, 10, Article 1621269. https://doi.org/10.3389/feduc.2025.1621269

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development.
Learning Policy Institute. https://doi.org/10.54300/122.311
Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. Educational Researcher, 38(3), 181–199. https://doi.org/10.3102/0013189X08331140
Erdogan, G., & Mede, E. (2021). The evaluation of an English preparatory program using CIPP model and exploring A1 level students' motivational beliefs. Journal of Education and Educational Development, 8(1), 53–76. https://doi.org/10.22555/joeed.v8i1.109
Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. American Educational Research Journal, 38(4), 915–945. https://doi.org/10.3102/00028312038004915
Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds. Teaching and Teacher Education, 68, 99–113. https://doi.org/10.1016/j.tate.2017.08.007
Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. Educational Assessment, Evaluation and Accountability, 26(1), 5–28. https://doi.org/10.1007/s11092-013-9179-5
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), Article 100607. https://doi.org/10.1016/j.ssaho.2023.100607
Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. Review of Educational Research, 88(4), 547–588. https://doi.org/10.3102/0034654318759268
Lippe, M., & Carter, P. (2018). Using the CIPP model to assess nursing education program quality and merit. Teaching and Learning in Nursing, 13(1), 9–13. https://doi.org/10.1016/j.teln.2017.09.008
Montero-Mesa, L., Fraga-Varela, F., Vila-Couñago, E., & Rodríguez-Groba, A. (2023). Digital technology and teacher professional development: Challenges and contradictions in compulsory education. Education Sciences, 13(10), Article 1029. https://doi.org/10.3390/educsci13101029
OECD. (2016). Supporting teacher professionalism: Insights from TALIS 2013. https://doi.org/10.1787/9789264248601-en
OECD. (2025). Unlocking high-quality teaching. https://doi.org/10.1787/f5b82176-en
Rahmi, I., & Rassanjani, S. (2025). Enhancing teacher quality in Indonesia: The impact of teacher professional development on achieving sustainable development goal 4.c. Social Sciences & Humanities Open, 12, Article 102123. https://doi.org/10.1016/j.ssaho.2025.102123
Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), International handbook of educational evaluation (pp. 31–62). Springer. https://doi.org/10.1007/978-94-010-0309-4_4
Stufflebeam, D. L., & Coryn, C. L. S. (2014). Evaluation theory, models, and applications (2nd ed.). Jossey-Bass.
UNESCO. (2024). Global education monitoring report 2024/5: Leadership in education—Lead for learning. https://unesdoc.unesco.org/ark:/48223/pf0000391406
Ventista, O. M., & Brown, C. (2023). Teachers' professional learning and its impact on students' learning outcomes: Findings from a systematic review. Social Sciences & Humanities Open, 8(1), Article 100565. https://doi.org/10.1016/j.ssaho.2023.10056