

Embedding Formative Evaluation in Qur'an Literacy: A CIPPO Model Study in a Public Secondary School

Asniar Tanjung*, Moh.Ali, A. Syathori, Agung

Institut Agama Islam Negeri Datuk Laksemana Bengkalis, Indonesia

Email Corresponding : asniartanjung@gmail.com

Abstract

Many students at general state secondary schools still cannot read or write the Qur'an correctly, yet the programs meant to address this weakness are seldom evaluated in any systematic way. This study evaluates the Qur'an literacy program at SMA Negeri 6 Konawe Selatan through the CIPPO model, which examines context, input, process, product, and output. The research used a qualitative descriptive-analytic design with a social phenomenological approach. Informants included the school principal, teachers, and students, chosen through purposive sampling. Data were gathered by interview, observation, and documentation, then analyzed through data reduction, data display, and conclusion drawing, while validity was secured through source, technique, and time triangulation together with member checking. The findings show that the program functioned well across all five components. Context evaluation confirmed a genuine need and a supporting module, input evaluation showed adequate resources and scheduling, and process evaluation recorded 70% of the monitored regular students reaching the minimum standard. Product evaluation verified sound attainment in reading and correct writing, and output evaluation yielded a mean score of 75.76 across 100 students. Notably, evaluation operated continuously within instruction rather than as a final audit, which allowed the teacher to correct weaknesses while learning was still underway. The study implies that schools can raise Qur'an literacy more effectively by embedding formative evaluation into everyday teaching, and it offers school leaders a practical basis for deciding whether to continue, revise, or strengthen such programs.

How to Cite:

Tanjung, A., Ali, M., Syathori, A., & Agung. (2025). Embedding Formative Evaluation in Qur'an Literacy: A CIPPO Model Study in a Public Secondary School. *Educazione: Journal of Education and Learning*, 3(1), 123-134.

Article History

Received : July 2025
Revised : August 2025
Accepted : November 2025

Keywords:

Qur'an Literacy Program, Educational Program Evaluation, Formative Evaluation, Character Formation

INTRODUCTION

The ability to read and write the Qur'an remains a foundational competence in Indonesian Islamic education, yet a persistent share of students in formal state schools still enters secondary level without mastering it (Mamlu'ah et al., 2025; Mustika Rini, 2024). This gap is not trivial, because Qur'an literacy underpins religious practice, character formation, and further study of Islamic knowledge, so its absence weakens the wider aims of Islamic Religious Education (Lestari et al., 2024). At SMA Negeri 6 Konawe Selatan the problem was concrete rather than abstract, since a number of students could not yet read the Qur'an correctly and some had not even mastered the *hijaiyah* letters, which prompted the school to run a structured reading and writing program with memorization of short verses as a condition for grade promotion. Programs of this kind, however, are frequently implemented without systematic evaluation, so schools rarely know whether the effort actually raises student competence (Li et al., 2024; Lynch & Prins, 2023). This situation makes a rigorous evaluation of such a program both timely and necessary.

A growing body of scholarship has applied the CIPP evaluation model to Qur'an-related programs in Islamic education, and three themes recur across these studies. The first theme

concerns program design and institutional support, where studies find that memorization programs succeed when they are anchored in a clear vision yet falter when external backing is weak; Muawanah et al. (2022) reported that a mandatory *tahfidz* program in *Yogyakarta madrasahs* ran well overall but suffered from limited parental support and product outcomes below target. The second theme addresses curriculum and character outcomes, illustrated by Mukhibat et al. (2024), who used a CIPP-informed case study to develop and evaluate a religious moderation curriculum in higher education and showed that evaluation can reshape both content and student attitudes. The third theme covers comparative implementation in boarding schools, where Ayyusufo et al. (2022) found *tahfiz* programs across two *pesantren* to be strong in context yet uneven in input and process. Together these works establish CIPP as a credible lens while concentrating on memorization in religious institutions.

The reviewed literature nonetheless leaves a clear gap that this study addresses. Prior CIPP evaluations cluster around *tahfidz* or memorization programs situated in *madrasahs*, *pesantren*, and *Islamic higher education*, and they rarely examine basic reading and writing literacy of the Qur'an within a general state secondary school (Duan et al., 2023; Irene, 2023). They also tend to treat the four classic CIPP components and to stop at product, without extending the analysis to an output stage that gauges the program's overall success against learning outcomes. This study departs from that pattern on three counts: its setting in a public senior high school rather than a religious institution, its focus on foundational reading and writing literacy rather than memorization alone, and its use of the extended CIPPO model that adds an explicit output dimension. The novelty therefore lies in bringing a five-stage evaluation to a program and a school type that the existing literature has largely overlooked.

The purpose of this study is accordingly to evaluate the Qur'an literacy program at SMA Negeri 6 Konawe Selatan using the CIPPO model, examining in turn its context, input, process, product, and output. By doing so the study aims to determine not only whether the program achieved its objectives but also how each evaluative stage functioned in the daily work of teaching. This purpose extends the existing literature in two directions, since it carries CIPP-based evaluation into the under-studied terrain of general secondary schooling and adds the output dimension that most prior studies omit. The relevance of the work is both academic and practical, because it offers researchers a fuller evaluative account of a literacy program while giving school leaders and teachers evidence on which to base decisions to continue, revise, or strengthen such programs.

The central focus of this study is to explore, in depth, how the CIPPO evaluation actually operated as an integral part of Qur'an literacy instruction rather than as an external audit imposed after the fact. The analysis attends to the lived practice of the program, drawing on what teachers and students reported, what the researcher observed in the classroom, and what the program's documents recorded, so that each finding rests on convergent evidence. Particular attention is paid to how the early stages informed later ones and how monitoring during the process shaped teaching while it was still underway. This qualitative orientation seeks understanding of the program's inner workings rather than a simple verdict on its success. The following section sets out the research method through which this focus was pursued.

RESEARCH METHOD

This study employed a qualitative design of the descriptive-analytic type, framed by a social phenomenological approach that seeks to understand a program as its participants experience and account for it (Urcia, 2021). The choice suited the aim of the study, which was to describe and interpret the real conditions of the Qur'an literacy program at SMA Negeri 6 Konawe Selatan rather than to test a hypothesis or measure an intervention. The site was selected purposively because a number of its students could not yet read the Qur'an correctly and some had not mastered the *hijaiyah* letters, which made it a fitting setting in which to examine how an evaluation of the program actually worked (Ahmad & Wilkins, 2025; Denieffe,

2020). Informants were determined through purposive sampling (Ahmad & Wilkins, 2025; Denieffe, 2020; Zheng & Thomas, 2025), that is, by choosing those who genuinely understood the internal and external conditions of the school and could therefore give accurate accounts of the program; they comprised the school principal, teachers, and students. This grounding in participants who lived the program gave the study the depth of insight that a phenomenological reading of the CIPPO evaluation required.

Data were gathered through three complementary techniques: interviews, observation, and documentation (Maxwell, 2022; Scheibelhofer, 2023; Seim, 2024). The interviews elicited how the principal, teachers, and students understood and evaluated the program across the context, input, process, product, and output components, providing the participants' own voices to anchor the findings. Observation allowed the researcher to observe instruction as it occurred, recording the opening routines, the letter-joining and memorization activities, and the assessment moments that the interviews could only describe from memory. *Documentation supplied* the material trace of the program, including the teaching module with its graded materials, tests, and assignments, and the semester test scores of the full body of 100 students. Combining these three techniques allowed the study to triangulate what was said, observed, and recorded, ensuring that claims confirmed across all three sources could be held with reasonable confidence (Lemon & Hayes, 2020; Mustafa, 2021; Rose & Johnson, 2020). Table 3 presents the mapping of these techniques to the CIPPO components to make their alignment explicit.

The correspondence between each data source and the evaluative component it chiefly informed is summarized in Table 1.

Table 1. Data Collection Techniques Mapped to the CIPPO Components

CIPPO component	Interview	Observation	Documentation
Context	Teachers' needs diagnosis and lesson purpose	Opening routine, review of prior material	Teaching module (materials, tests, assignments)
Input	Sourcing, scheduling, and grouping decisions	Preparation of materials before the lesson	Program schedule and lesson design
Process	Monitoring logic and obstacles faced	Letter-joining and memorization activity	Daily-test records of the regular group
Product	Judgment of goal attainment	Drills and mid/end-semester testing	Assessed student work across classes X–XII
Output	Overall success in reading the Qur'an	Assessment of learning outcomes	Semester test scores (n = 100)

Source: Developed by the researcher at SMA Negeri 6 Konawe Selatan (2022).

Table 1 shows that no single component rested on one technique alone, since every stage of the CIPPO evaluation was approached from at least three angles. This deliberate overlap is what turned the collection design into a triangulation instrument rather than a simple checklist, because a finding under any component could be cross-read against an interview, an observation, and a document. The mapping also clarifies why the documentary sources carried decisive weight at the output stage, where the semester scores of one hundred students gave the fullest picture of overall success, while the interviews carried more of the interpretive load at the context stage. Read as a whole, the table demonstrates that the study's confidence in each finding grew from the convergence of distinct sources rather than from the volume of data in any one of them.

Data were analyzed using the interactive model of data reduction, data display, and conclusion drawing and verification, which was carried out as a continuous cycle rather than a fixed sequence (Laari, 2025). In the reduction stage, raw interview transcripts, field notes, and documents were sorted and sharpened around the five CIPPO components, setting aside irrelevant material and highlighting significant patterns. In the display stage, the reduced data were arranged into organized accounts and tables under each component, making the relationships among context, input, process, product, and output visible and ready for

interpretation. In the final stage, the researcher drew conclusions from these displays and verified them, testing each emerging claim against the body of evidence before treating it as a finding. Data validity was secured through triangulation of sources, techniques, and time, as well as member checking, ensuring that the reported conclusions rest on evidence confirmed from several directions and verified by informants (Farquhar et al., 2020).

RESULT AND DISCUSSION

Results

Evaluation of the Qur'an Literacy Program Using the CIPPO Model

Learning stands at the core of the whole educational process, and the teacher acts as its principal agent. Teacher and students together carry out the aims of instruction, and those aims reach their fullest result only when teaching and learning run effectively. Within this program, the CIPPO model organizes the evaluation into five connected components: context, input, process, product, and output. The findings below take each component in turn and draw on three data sources that were kept distinct at every stage: interviews with informants, the researcher's direct classroom observation, and the program's teaching documents. To orient the reader before the detailed findings, the sequence and guiding question of each evaluative stage is set out in Figure 1.

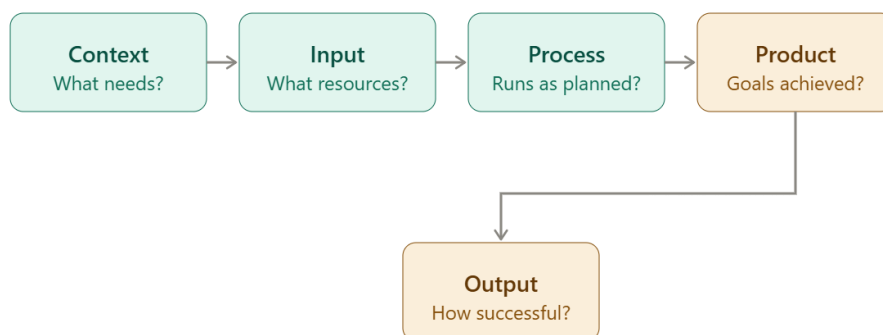


Figure 1. The CIPPO evaluation cycle applied to the BTQ program

Source: Processed from the study's field data, SMA Negeri 6 Konawe Selatan, 2022.

Figure 1 clarifies that the five stages form a single working cycle rather than a set of separate checks, since each stage answers a different question yet feeds the one after it. The context stage identifies what the program must address, the input stage assembles the means to address it, and the process stage watches those means in action. The product and output stages then close the loop by asking, in turn, whether the stated objectives were met and how far the program succeeded as a whole. Reading the findings against this map helps explain why a figure reported at one stage, such as the mid-process attainment of a single group, need not match a figure reported at another, such as the final result of the entire cohort.

Context Evaluation

Context evaluation examined the situational grounds on which the instruction was planned, that is, the teacher's reading of student needs and the preparatory decisions taken before teaching began. The interview data established how teachers framed the purpose of the program, as one informant explained: *"Evaluating the Qur'an literacy program by attending to its context means the teacher helps plan decisions and shows the needs to be achieved by the Qur'an literacy program. For example, before teaching, the preparation the teacher makes is to develop a module as a guide for teaching BTQ; inside the module there are materials, tests, and assignments."* (Adil, Interview, 27 May 2022)

A second informant, interviewed separately, described the same contextual grounding but stressed the classroom routine through which student understanding was gauged: *"The Qur'an*

literacy program, attending to its context, means the teacher helps plan the decisions and shows the needs to be achieved by the program. In the classroom, what the teacher does on entering is to open with a greeting that the students answer in unison; on this occasion the teacher takes attendance and reviews the previous material, namely writing sentences. This context evaluation needs to be done to see how far the students understand and as reflective material for the teacher." (Ay Romlah, Interview, 29 May 2022)

Observation corroborated these accounts at the level of practice rather than repeating them, and the field record kept the observer's own words intact: *"In the observation conducted on 20 May 2022, 'The teacher helps plan decisions and shows the needs to be achieved by the Qur'an literacy program. For example, before teaching, the preparation made is to develop a module as a guide for teaching BTQ; inside the module there are materials, tests, and assignments. In Qur'an literacy learning, what the teacher does on entering the class is to open with a greeting that the students answer in unison; on this occasion the teacher takes attendance and reviews the previous material, namely writing sentences. This context evaluation needs to be done to see how far the students understand and as reflective material for the teacher.'"* (Field observation, 20 June 2022)

In the core activity, the researcher observed the teacher explaining changes in letter form, covering isolated letters and letters joined at the start, middle, and end of a word, and then writing those forms on the board. Most students were attentive and the majority asked questions actively, while a smaller number stayed silent yet still followed the explanation. Document analysis completed the triangulation, since the module examined as an artifact was built around graded materials, tests, and assignments. That artifact confirmed that the planning the informants described was written into the program's documents rather than merely intended.

The three sources converge on one finding while illuminating different sides of it. The interviews reveal the teacher's intent, treating context evaluation as a needs diagnosis and a reflective device, and the observation shows that intent enacted in a structured opening, review, and boardwork that let the teacher read comprehension in real time. The module then supplies the material trace that anchors both accounts. This convergence matters because context evaluation is easily reduced to a paper exercise, yet here it worked as a live diagnostic that required no new curriculum before evaluation could begin. The teacher used the module together with the classroom review to locate the gap between students' actual ability and the program's target, a gap felt most sharply among rural learners whose economic circumstances were broadly similar but whose baseline in Qur'an reading and writing stayed low. On this basis the program treated the memorization of several short verses and their correct writing as an added strand within BTQ and a condition for grade promotion.

Input Evaluation

Input evaluation concerns the resources, schedule, and working procedures the teacher implemented so that the program's aims could be pursued. The observation record supplied the first evidence, and the observer's notes is preserved in full: *"In the observation conducted on 20 May 2022, the teacher helps arrange decisions, for example by determining the available sources (by asking students to reproduce the teaching material) and setting the schedule on which the Qur'an literacy program will be carried out. In addition, the teacher holds an evaluation (practice questions or practicals) for the students, and further determines the working procedure, for example by using small groups and preparing everything needed when the lesson is about to begin."* (Field observation, 20 June 2022)

The interview data matched this account and allowed two informants to address complementary sides of the same practice. The first stressed the sourcing and scheduling decisions: *"Evaluating the Qur'an literacy program by attending to input means the teacher helps arrange decisions, for example by determining the available sources (by asking students to reproduce the teaching material) and setting the schedule on which the Qur'an literacy program will be carried out."* (Daniyanti, Interview, 03 June 2022)

The second informant, asked separately, foregrounded assessment and grouping as part of the same input stage: *"The Qur'an literacy program, attending to input, means the teacher holds an evaluation (practice questions or practicals) for the students. In addition, the teacher determines the working*

procedure, for example by using small groups and preparing everything needed when the lesson is about to begin." (Harmina, Interview, 05 June 2022)

When read across the three sources, the finding is that input evaluation here turned mainly on deciding what resources and strategies to deploy and how to organize them. Observations show the arrangement in action, the two interviews identify its two components, and repeated references to the reproduced teaching materials links input back to the module already documented under context. One resource stood out as decisive, namely the availability of a competent teacher. In a program of this kind, an able instructor is not one input among many, but the condition on which the others depend, and the data indicate that the school's capacity to run the memorization-and-writing requirement rested on having such a teacher in place. Input evaluation therefore worked as a readiness check, confirming that materials, timetable, grouping, and teaching capacity were assembled before the process stage began.

Table 2. Implementation Design of the *BTQ* Program Process at SMA Negeri 6 Konawe Selatan

Component	Description of the field activity
Schedule	Morning <i>tilawah</i> is held daily for 15 minutes before lessons begin, and memorization is carried out during each lesson period.
Target	Graduating students are able to read the Qur'an and to memorize and correctly write several short verses.
Method	Correct-writing and memorization drills are conducted during <i>BTQ</i> lessons and outside <i>BTQ</i> hours.
Evaluation	Assessment is embedded in every <i>BTQ</i> session, covering the initial, mid, and end stages of the activity.

Source: Processed from field observation and program documents at SMA Negeri 6 Konawe Selatan, 2022.

Table 2 shows that the process rested on a fixed rhythm rather than improvisation, built from daily *tilawah*, lesson-embedded memorization, a stated graduation target, and evaluation woven into each session. This design helps explain the 70% KKM attainment reported for the 35 regular students observed at this stage, since a predictable schedule and continuous assessment gave the teacher repeated points at which to detect weakness and correct it. The 70% and 30% figures should be read for what they are, namely a snapshot of one regular group monitored while the process was underway rather than the final standing of every program participant, which the output stage reports separately. The obstacle the teacher named, the diversity of students arriving from different regions with uneven starting ability, is exactly the weakness the process design was meant to surface early, and the assessment recorded here ran from an oral pretest given at random, through a mid-activity check of engagement that informed the final grade, to a practical test of reading and writing skill. Process evaluation thus operated as the program's self-correcting mechanism.

Product Evaluation

Product evaluation focused on measuring whether participants attained the objectives the program had set. The observation record stated the measurement logic and preserved the observer's phrasing: *"In the observation conducted on 20 May 2022, the teacher carries out assessments to measure success, namely by holding drills and tests, both mid-term and end-of-semester. Product evaluation is the final stage, that is, an evaluation of whether participants achieve the objectives that have been set. This is done to help those responsible for the program in deciding to modify or to stop it. This evaluation is carried out by the assessor in measuring the achievement of the objectives that were developed and administered, and its benchmark is the results of the children's tests or daily exercises, whether oral or written."* (Field observation, 20 June 2022)

Two informants confirmed this orientation. The first defined the product stage as a test of goal attainment: *"Evaluating the Qur'an literacy program by focusing on the product means the teacher carries out assessments to measure success, namely by holding drills and tests, both mid-term and end-of-semester. Product evaluation is an evaluation of whether participants achieve the objectives that have been set."* (Nur Arfah Lasao, Interview, 17 June 2022)

The second informant located the same judgment in the concrete record of student work: *"The Qur'an literacy program, focusing on the product, means the teacher carries out assessments. This evaluation is done by the assessor in measuring the achievement of the objectives that were developed and administered, and its benchmark is the results of the children's tests or daily exercises, whether oral or written."* (Nursia, Interview, 19 June 2022)

Across the three sources the finding holds steady, since product evaluation judged attainment against the program's own objectives using the gathered evidence of drills, mid-term tests, and semester examinations. The documentary record of student work supported the informants' claim that, for classes X, XI, and XII, attainment in the memorization strand was good, with students in the "good" category able to memorize the short verses and write them correctly in line with the rules of *tajwid*. The observation and the interviews agree that this outcome was not automatic, because both note that continued guidance from the BTQ teacher and the wider staff remained necessary to sustain it. Product evaluation therefore did more than certify results, since it fed a decision by giving those responsible the grounds to continue, modify, or halt the program on the strength of what the assessed work actually showed. In this sense the stage linked measurement directly to program governance.

Output Evaluation

Output evaluation asked, finally, how successful the program was as a whole once students had completed it. The observation record framed the question in the observer's own terms: *"In the observation conducted on 20 May 2022, the teacher conducted an assessment of the students' level of success. This evaluation was carried out by the assessor to measure the level of success of the objectives, and its benchmark was the students' learning outcomes."* (Field observation, 20 June 2022)

The interview data restated this as the closing stage of the whole evaluative cycle: *"Evaluating the Qur'an literacy program by attending to output means the teacher assesses the students' level of success in reading the Qur'an, that is, whether they can read the Qur'an after taking part in the learning. This output evaluation is the final stage, an evaluation of the program's level of success."* (Nur Arfah Lasao, Interview, 17 June 2022)

The documentary source carried the decisive weight at this stage, in the form of the semester test scores for the full body of 100 program participants. Because listing individual names would add nothing analytically and would compromise the confidentiality owed to students in a single-site study, the raw list is presented as a distribution in Table 3.

Table 3. Distribution of BTQ Semester Test Scores

Score band	Number of students	Percentage
90–100	12	12%
80–89	24	24%
70–79	28	28%
60–69	36	36%
Total	100	100%
Total score	7,576	
Mean score	75.76	

Source: BTQ semester test document, SMA Negeri 6 Konawe Selatan, 2022.

The same distribution is shown as a bar chart in Figure 2, which makes the shape of the results easier to read than a column of figures.

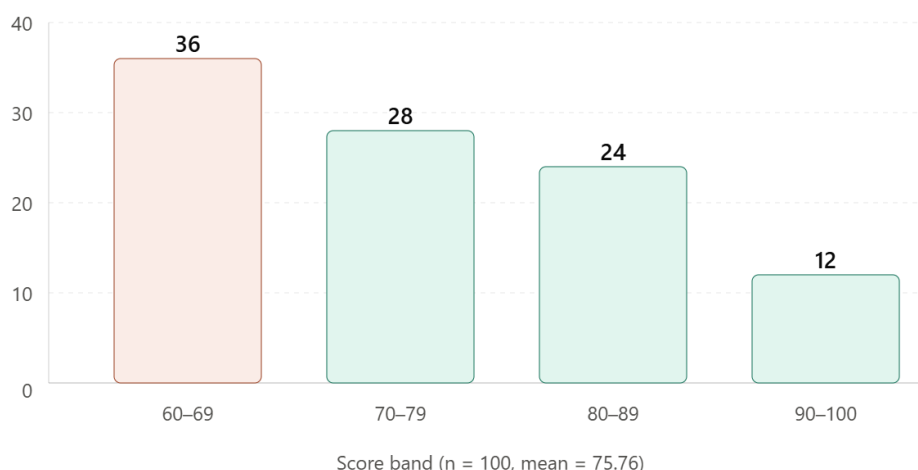


Figure 2. Bar chart of *BTQ* semester test score distribution

Source: Processed from the *BTQ* semester test document, SMA Negeri 6 Konawe Selatan, 2022.

Table 3 and Figure 2 together show a program whose overall output can fairly be called moderately successful, with a mean of 75.76 across one hundred students and a total of 7,576. The upper bands are far from empty, since 36 students scored 80 or above and 12 reached the 90-100 range, which indicates that a meaningful share of learners performed strongly. The shape of the distribution repays attention, because the largest single band sits at 60-69, just beneath the range where the earlier process figures placed the passing threshold, so much of the lower group clusters close to the line rather than falling far short. This output reading and the 70% reported under process describe different units rather than competing claims, since the 70% covered the 35 regular students monitored while the process was still running, whereas the 75.76 mean summarizes the final semester result of the full cohort. Placed side by side, the two readings tell one coherent story of a program that met its target for most participants while leaving a well-defined minority to be raised over time.

Discussion

The context and input findings, read together, confirm what CIPPO theory anticipates about the front end of a program yet add a detail worth dwelling on. Stufflebeam designed context evaluation to identify unmet needs and the conditions a program must address before it begins, and input evaluation to weigh the resources and strategies available to meet those needs (Birgili & Kırkıç, 2021; Huang, 2025). In this school, the teacher's module, the diagnostic classroom review, and the deliberate arrangement of schedule and grouping matched that design closely. What stands out is that the teacher did not treat these two stages as paperwork completed once and filed away. The module functioned as a living instrument that carried the needs diagnosis into every lesson, which suggests that context and input, in practice, were not a preliminary phase so much as a stance the teacher maintained throughout. This reading extends the standard account rather than contradicting it, since it locates the value of the early stages in their continuing use.

The process findings speak directly to the monitoring logic that Worthen and Sanders, as cited in Arikunto, place at the heart of program evaluation, namely the checking of whether an activity runs as planned so that weaknesses can be corrected while there is still time to act (Cimarolli et al., 2025; Mark, 2024). The 70% *KKM* attainment among the 35 regular students should be interpreted through this lens, because it was never meant as a final verdict but as a mid-course reading that told the teacher where to intervene. The teacher's own account of the obstacle, the uneven starting ability of students drawn from different regions, shows the process

stage doing exactly what it is designed to do, that is, surfacing a problem early enough to shape the response. Liu and Huang's (2020) argument that evaluation should improve the quality of learning rather than merely record it fits this pattern well. The daily *tilawah*, the lesson-embedded drills, and the assessment folded into each session gave the program repeated correction points rather than a single terminal test.

The product and output findings clarify a distinction that program evaluators often blur, and the data make the distinction concrete. Product evaluation, in Stufflebeam's framework, asks whether participants achieved the stated objectives and feeds the decision to continue, revise, or halt a program, while output evaluation gauges the overall level of success against learning outcomes (Lynch & Prins, 2023). In this study, the product stage certified that students across classes X, XI, and XII could memorize and correctly write the short verses in line with *tajwid*, and the output stage summarized the whole cohort with a mean of 75.76 across one hundred students. The two figures reported at different stages, the 70% among 35 monitored learners and the 64 of 100 who met the threshold at the end, were not in tension once their units were separated. Reading them side by side turned an apparent inconsistency into a coherent narrative of a program that succeeded for most participants while leaving a defined minority near the line.

Taken as a whole, these findings invite a rethinking of how the CIPPO sequence actually behaved in this setting. The textbook image of CIPPO is a straight line running from context through to output, with evaluation arriving at the end to pass judgment (Maharani & Hadiapurwa, 2023). The evidence here points to something different in character, because evaluation was distributed across the program rather than reserved for its conclusion. Assessment lived inside every session, monitoring during the process reshaped teaching in real time, and the early stages kept feeding the later ones through the module and the review routine. The program did not run the five stages once and stop; it cycled through diagnosis, action, and correction continuously, with each pass tightening the fit between what students could do and what the program required. This behavior is closer to a formative loop than to a linear audit.

To make that synthesis visible, the model that emerges from the findings is presented in Figure 3. It reframes the five CIPPO components not as a one-way pipeline but as an embedded formative cycle in which a continuous evaluative core touches every stage, and in which the product and output results loop back to re-inform context and input for the next teaching period. The novelty of this study lies precisely in this reframing, as it shows CIPPO operating as a self-correcting mechanism woven into instruction rather than as an external checklist applied after the fact. Before turning to the model, it is worth stating why this matters beyond the single site: a program that evaluates continuously builds its own feedback into the act of teaching, which is the condition under which evaluation can raise quality rather than merely certify it (Teig & Nilsen, 2022).

The embedded formative evaluation model derived from the CIPPO components at SMA Negeri 6 Konawe Selatan is presented in Figure 3. Methodologically, this study offers an embedded formative evaluation model that repositions CIPPO as a continuous, self-correcting cycle woven into Qur'an literacy instruction, a departure from the linear, terminal use of the model common in the literature (Maharani & Hadiapurwa, 2023). Second, in practical terms, it shows school leaders and *BTQ* teachers a concrete way to build correction points into ordinary lessons, through daily *tilawah*, session-level assessment, and mid-process monitoring, so that a diverse intake can be brought toward a shared standard over time. This continuous, reflective use of evaluation also opens a bridge toward transformational education and learning, because an evaluation that constantly returns information to both teacher and learner creates the reflective conditions under which character formation and durable religious literacy can take root (Bez et al., 2025). The present data document the mechanism rather than the transformation itself, so this bridge is best read as a direction for further study, in which the long-term formation of learners under an embedded formative model becomes the object of inquiry.

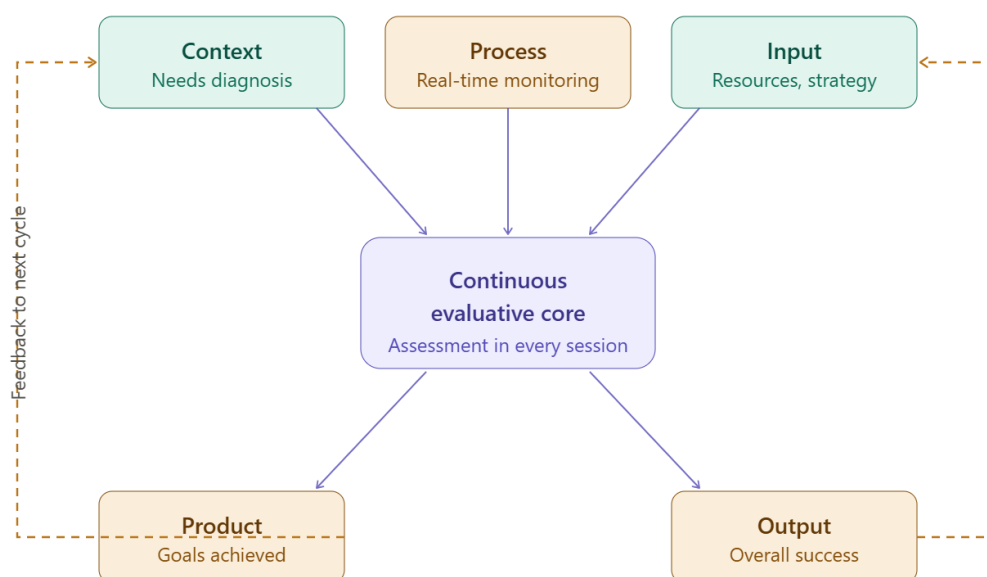


Figure 3. The embedded formative evaluation model derived from the CIPPO findings

CONCLUSION

The evaluation of the Qur'an literacy program at SMA Negeri 6 Konawe Selatan using the CIPPO model shows that the program ran effectively across its five components, with a needs-based context, well-arranged inputs, a monitored process, certified product attainment, and an overall output mean of 75.76 across one hundred students. The clearest pattern was that evaluation operated continuously within instruction rather than only at its close, allowing the teacher to detect and correct weaknesses while learning was still underway. This study nonetheless carries several limitations that temper its claims. It rests on a single school, relies on informant accounts alongside observation and documents that reflect one program cycle, and measures attainment rather than the longer-term formation of learners, so its findings should be read as context-bound rather than broadly generalizable. Future research could therefore test the embedded formative model across multiple schools, follow students over several cycles to observe durable outcomes, and combine the qualitative account with wider quantitative measures so that the model's transferability and lasting effects can be examined with greater confidence.

ACKNOWLEDGMENT

The author thanks the principal, teachers, and students of SMA Negeri 6 Konawe Selatan for their willingness to take part in this research and for the access granted during the fieldwork. Appreciation is also extended to colleagues who offered valuable comments on earlier drafts of this article.

REFERENCES

- Ahmad, M., & Wilkins, S. (2025). Purposive Sampling in Qualitative Research: A Framework for the Entire Journey. *Quality and Quantity*, 59(2), 1461–1479. <https://doi.org/10.1007/s11135-024-02022-5>

- Ayyusufi, A. M., Anshori, A., & Muthoifin, M. (2022). Evaluation of The CIPP Model on The Tahfidz Program in Islamic Boarding Schools. *Nazhruna: Jurnal Pendidikan Islam*, 5(2), 466–484. <https://doi.org/10.31538/nzh.v5i2.2230>
- Bez, S., Burkart, F., Tomasik, M. J., & Merk, S. (2025). How Do Teachers Process Technology-Based Formative Assessment Results in Their Daily Practice? Results From Process Mining of Think-Aloud Data. *Learning and Instruction*, 97, 102100. <https://doi.org/10.1016/j.learninstruc.2025.102100>
- Birgili, B., & Kırkıç, K. A. (2021). Evaluation of a Strategic Management Program: Context, Input, Process, Product Model as a Prototype for Business Academies. *TEM Journal*, 10(1), 204–214. <https://doi.org/10.18421/TEM101-26>
- Cimarolli, V., Sanders, A., Wylie, M., & Stone, R. (2025). Evaluation of a Wellness Program For Affordable Senior Housing Residents. *Innovation in Aging*, 9(Supplement_2). <https://doi.org/10.1093/geroni/igaf122.3002>
- Denieffe, S. (2020). Commentary: Purposive Sampling: Complex or Simple? Research Case Examples. *Journal of Research in Nursing*, 25(8), 662–663. <https://doi.org/10.1177/1744987120928156>
- 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). <https://doi.org/10.1177/21582440221149415>
- Farquhar, J., Michels, N., & Robson, J. (2020). Triangulation in Industrial Qualitative Case Study Research: Widening the Scope. *Industrial Marketing Management*, 87, 160–170. <https://doi.org/10.1016/j.indmarman.2020.02.001>
- Huang, S. H. (2025). Evaluating the English for General Purposes (EGP) Program at a Taiwanese University: A CIPP (Context, Input, Process, and Product) Model Study. *Evaluation and Program Planning*, 112, 102662. <https://doi.org/10.1016/j.evalprogplan.2025.102662>
- 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 and Humanities Open*, 8(1), 100607. <https://doi.org/10.1016/j.ssaho.2023.100607>
- Laari, L. (2025). Inductive-Deductive Qualitative Data Analysis Logic in Health Sciences Research: A Framework for Analysing Qualitative Data. *International Journal of Qualitative Methods*, 24. <https://doi.org/10.1177/16094069251381706>
- Lemon, L. L., & Hayes, J. (2020). Enhancing Trustworthiness of Qualitative Findings: Using Leximancer for Qualitative Data Analysis Triangulation. *Qualitative Report*, 25(3), 604–614. <https://doi.org/10.46743/2160-3715/2020.4222>
- Lestari, P., Ghufron, A., & Surjono, H. D. (2024). Teachers' Experiences in Implementing the School Literacy Movement (GLS): An Investigation of Junior High School Literacy Programs in Indonesia. *Journal of Curriculum and Teaching*, 13(5), 393–404. <https://doi.org/10.5430/jct.v13n5p393>
- Li, A., Miller, F. G., & Williams, S. C. (2024). Cultural Adaptations to Social-Emotional Learning Programs: A Systematic Review. *School Psychology*, 40(2), 108–120. <https://doi.org/10.1037/spq0000649>
- Liu, Y., & Huang, J. (2020). The Quality Assurance of a National English Writing Assessment: Policy Implications for Quality Improvement. *Studies in Educational Evaluation*, 67, 100941. <https://doi.org/10.1016/j.stueduc.2020.100941>
- Lynch, J., & Prins, E. (2023). Knowledge to Support Early Educators and Elementary School Teachers' Engagement in Family Literacy and Family Literacy Programs. *Reading Teacher*, 76(5), 618–626. <https://doi.org/10.1002/trtr.2187>
- Maharani, A. P., & Hadiapurwa, A. (2023). Evaluation of High School Library Management: Implementation of CIPPO Evaluation Model. *Record and Library Journal*, 9(1), 118–126. <https://doi.org/10.20473/rlj.V9-I1.2023.118-126>

- Mamlu'ah, A., Rahmawati, U. B., & Jalil, A. (2025). Development of Tahsin Tahfidz Media Innovation "Memo Qur'an" in Improving Reading Literacy and Digital Literacy of Students. *Ablussunnah: Journal of Islamic Education*, 4(3), 224–235. <https://doi.org/10.58485/jie.v4i3.396>
- Mark, M. M. (2024). Blaine Worthen: Good Steward of AJE and More. *American Journal of Evaluation*, 45(1), 158–159. <https://doi.org/10.1177/10982140231225235>
- Maxwell, J. A. (2022). Response to David Morgan on Triangulation. *Journal of Mixed Methods Research*, 16(4), 412–414. <https://doi.org/10.1177/15586898221122758>
- Muawanah, S., Said, A., Furqoni, R., Muzayanah, U., & Mustolehudin. (2022). Evaluating Mandatory Tahfīz Quran Program Implementation At Madrasah. *Jurnal Pendidikan Islam*, 8(2), 239–254. <https://doi.org/10.15575/jpi.v8i2.20330>
- Mukhibat, M., Effendi, M., Setyawan, W. H., & Sutoyo, M. (2024). Development and Evaluation of Religious Moderation Education Curriculum at Higher Education in Indonesia. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2302308>
- Mustafa, C. (2021). Qualitative Method Used in Researching the Judiciary: Quality Assurance Steps to Enhance the Validity and Reliability of the Findings. *Qualitative Report*, 26(1), 176–185. <https://doi.org/10.46743/2160-3715/2021.4319>
- Mustika Rini, A. W. (2024). Improving Al-Qur'an Literacy: CTL Model, Digital Al-Qur'an, and Short Surah Reading Skills in PAI Learning. *Journal of Education Technology and Inovation*, 7(1), 79–86. <https://doi.org/10.31537/jeti.v7i1.1870>
- Rose, J., & Johnson, C. W. (2020). Contextualizing Reliability and Validity in Qualitative Research: Toward More Rigorous and Trustworthy Qualitative Social Science in Leisure Research. *Journal of Leisure Research*, 51(4), 432–451. <https://doi.org/10.1080/00222216.2020.1722042>
- Scheibelhofer, E. (2023). The Interpretive Interview: An Interview Form Centring on Research Participants' Constructions. *International Journal of Qualitative Methods*, 22. <https://doi.org/10.1177/16094069231168748>
- Seim, J. (2024). Participant Observation, Observant Participation, and Hybrid Ethnography. *Sociological Methods and Research*, 53(1), 121–152. <https://doi.org/10.1177/0049124120986209>
- Teig, N., & Nilsen, T. (2022). Profiles of Instructional Quality in Primary and Secondary Education: Patterns, Predictors, and Relations to Student Achievement and Motivation in Science. *Studies in Educational Evaluation*, 74, 101170. <https://doi.org/10.1016/j.stueduc.2022.101170>
- Urcia, I. A. (2021). Comparisons of Adaptations in Grounded Theory and Phenomenology: Selecting the Specific Qualitative Research Methodology. *International Journal of Qualitative Methods*, 20, 16094069211045474. <https://doi.org/10.1177/16094069211045474>
- Zheng, H., & Thomas, S. M. (2025). Exploring Chinese Stakeholders' Perceptions of School Inspection Criteria in Demonstrating Educational Quality: Do Discrepancies in Opinions Reflect Different School Contexts? *Studies in Educational Evaluation*, 86, 101481. <https://doi.org/10.1016/j.stueduc.2025.101481>