Kuttab : Jurnal Ilmu Pendidikan Islam
p-ISSN (2549-7987), e-ISSN (2550-0341)
Vol. 10, No. 2, 2026, pp. 318-332

Context and Input Evaluation of Arabic Language Learning in *Pesantren*: A CIPP Model Study at Tarbiyatut Tholabah Foundation

Ahmad Nahrowi[1*], Durrotun Aniqoh[2], Abdul Manan[3]
Universitas Tarbiyatut Tholabah Lamongan
ahmadrowi95@gmail.com, d.aniqoh@gmail.com, aabd.manan@gmail.com
*Corresponding author

DOI: 10.30736/ktb.v10i2.2945

Article History
Received : 22-07-2026
Revised : 23-09-2026
Accepted : 24-09-2026

Keywords: Arabic Language Learning, CIPP Model, *Pesantren*, Program Evaluation

Abstract: Learning Arabic is strategically important in *pesantren*-based education because it facilitates access to Islamic knowledge and fosters language competence. This study evaluated the Context and Input components of the CIPP (Context, Input, Process, Product) framework in the Arabic-language learning program at the Tarbiyatut Tholabah in Kranji, Paciran, Lamongan. A convergent parallel mixed-methods design was employed by collecting quantitative and qualitative data concurrently, analyzing them independently, and integrating the findings during interpretation. The Context evaluation showed that the program was aligned with the foundation's educational vision but lacked a formal evidence-based needs assessment and an explicit framework for aligning and progressing Arabic competencies across educational levels. The Input evaluation produced an overall score of 59.1%, categorized as moderately adequate. Curriculum and instructional documents achieved the highest score at 71.0%, followed by infrastructure at 62.5%, instructional media and teaching materials at 58.0%, and teacher qualification alignment at 45.0%. These findings reveal a gap between strong institutional commitment and the readiness of contextual planning and program resources.

Kata kunci:
Pembelajaran Akidah Akhlak, Pendidikan Islam, Regresi Linear Sederhana, Perilaku Siswa, Karakter Siswa

Abstrak: *Pembelajaran bahasa Arab memiliki arti strategis dalam pendidikan berbasis pesantren karena memfasilitasi akses terhadap pengetahuan keislaman dan mengembangkan kompetensi berbahasa. Penelitian ini mengevaluasi komponen Context dan Input dalam kerangka CIPP (Context, Input, Process, Product) pada program pembelajaran bahasa Arab di Tarbiyatut Tholabah, Kranji, Paciran, Lamongan. Desain metode campuran paralel konvergen digunakan dengan mengumpulkan data kuantitatif dan kualitatif secara bersamaan, menganalisisnya secara terpisah, dan mengintegrasikan temuan pada tahap interpretasi. Evaluasi Context menunjukkan bahwa program tersebut selaras dengan visi pendidikan yayasan, tetapi belum memiliki asesmen kebutuhan berbasis bukti yang formal serta kerangka eksplisit untuk menyelaraskan dan mengembangkan kompetensi bahasa Arab secara bertahap di berbagai jenjang pendidikan. Evaluasi Input menghasilkan skor keseluruhan sebesar 59,1%, yang dikategorikan cukup memadai. Kurikulum dan dokumen pembelajaran memperoleh*

Copyright © 2026 by Ahmad Nahrowi, Durrotun Aniqoh, Abdul Manan. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

skor tertinggi sebesar 71,0%, diikuti oleh sarana dan prasarana sebesar 62,5%, media pembelajaran dan bahan ajar sebesar 58,0%, serta kesesuaian kualifikasi guru sebesar 45,0%. Temuan ini mengungkapkan adanya kesenjangan antara komitmen kelembagaan yang kuat dan kesiapan perencanaan kontekstual serta sumber daya program.

INTRODUCTION

Arabic language learning occupies a strategic position within *pesantren*-based Islamic educational institutions because it serves not only as a medium for developing foreign-language competence but also as a primary instrument for accessing and understanding Islamic sources, including the Qur'an, Hadith, and classical Islamic literature. Consequently, the quality of Arabic language instruction has become an important component in achieving the broader educational objectives of *pesantren*. As educational institutions continue to respond to changing academic, social, and technological demands, Arabic language programs are expected to be implemented through systematic planning, adequate resources, and continuous quality improvement mechanisms to ensure sustainable educational outcomes (Muhajir, 2021; Tauhid, 2026).

Despite its importance, Arabic language learning in many *pesantren* continues to face persistent challenges. Previous studies have reported recurring issues related to curriculum alignment, instructional approaches, teacher competence, learning resources, and institutional support systems, all of which may affect the effectiveness of teaching and learning processes (Akhyar et al., 2025; Hadisi & Muna, 2025; Sapura et al., 2025). Preliminary observations conducted at the Tarbiyatut Tholabah in Kranji, Paciran, Lamongan, similarly indicated concerns about variations in learning achievement across educational levels, limitations in learning media and instructional resources, and disparities in teachers' academic qualifications. These conditions suggest the need for a systematic evaluation to identify both contextual and institutional factors that may influence the readiness and quality of Arabic language learning programs.

Program evaluation has been widely recognized as an essential mechanism for supporting educational quality improvement and evidence-based decision making. Among various evaluation approaches, the Context, Input, Process, Product (CIPP) model, developed by Stufflebeam, is one of the most widely used frameworks because it enables a systematic assessment of program planning, resource availability, implementation conditions, and outcomes. In educational settings, the CIPP model has been extensively used to evaluate curriculum implementation, instructional programs, institutional quality assurance systems, and learning effectiveness. Previous studies have demonstrated that contextual conditions and program inputs are important determinants of the quality of educational implementation and may contribute to achieving intended educational outcomes (Abdilah, 2025; Adlya et al., 2025; Arofah, 2021; Fitri et al., 2026).

However, a review of the existing literature reveals that most CIPP-based studies in Arabic language education have either evaluated programs comprehensively across all four dimensions or focused primarily on learning outcomes and curriculum

implementation. Limited attention has been given to an in-depth examination of the Context and Input dimensions as foundational components that determine program readiness prior to implementation. Furthermore, studies investigating these dimensions within an integrated *pesantren* educational ecosystem that encompasses multiple educational units under a single institutional foundation remain relatively scarce. As a result, empirical evidence regarding the extent to which contextual conditions, human resources, facilities, learning media, and institutional support collectively shape the readiness of Arabic language learning programs in *pesantren* settings remains limited.

The novelty of this study lies in its focused evaluation of the Context and Input dimensions of the CIPP framework within an integrated *pesantren* educational environment consisting of both formal and non-formal educational units under the Tarbiyatut Tholabah. Rather than conducting a broad evaluation of all CIPP dimensions, this study provides a more detailed diagnostic analysis of program relevance, institutional readiness, teacher qualifications, learning facilities, and supporting resources as fundamental prerequisites for effective program implementation. Such an approach is expected to generate more targeted evidence for institutional improvement and policy formulation.

Therefore, this study aims to evaluate the Context and Input dimensions of the Arabic language learning program at the Tarbiyatut Tholabah, Kranji, Paciran, Lamongan, using the CIPP evaluation framework. Specifically, the study seeks to answer the following research questions: To what extent do the contextual conditions support the implementation of Arabic language learning programs at the Tarbiyatut Tholabah Foundation? and to what extent are the available inputs, including human resources, learning facilities, instructional media, and institutional support, adequate to support effective Arabic language learning?

METHOD

This study employed a convergent parallel mixed-methods design in which quantitative and qualitative data were collected during the same research period, analyzed independently, and integrated during the interpretation stage. Quantitative data were used to assess the adequacy of program inputs. In contrast, qualitative data were used to explain institutional needs, program gaps, policies, and resource conditions that underlie the quantitative results (Creswell & Clark, 2017).

The study adopted the Context, Input, Process, Product (CIPP) evaluation framework but deliberately limited its scope to the Context and Input components. This limitation reflected the evaluation's diagnostic orientation, which focused on identifying program needs and assessing institutional readiness before examining program implementation and outcomes. Context evaluation examined institutional alignment, needs identification, and competency progression, while Input evaluation assessed teacher readiness, instructional resources, curriculum documents, and learning-support facilities (Stufflebeam & Zhang, 2017). Accordingly, this study did not evaluate classroom implementation as a Process component or student learning outcomes as a Product component.

The research was conducted at the Tarbiyatut Tholabah Foundation, Kranji, Paciran, Lamongan, East Java, Indonesia, from January to March 2026. The institution was purposively selected because it administers an integrated *pesantren*-based educational system comprising Qur'anic education, Madrasah Diniyah, and formal educational units with different Arabic language learning orientations. The primary unit of analysis was the Arabic language learning program administered by the foundation. Institutional leaders, teachers, curriculum documents, instructional resources, and facilities were treated as sources of evidence representing the evaluated program.

Participants in the qualitative component were selected through purposive sampling based on their involvement in program planning, coordination, and implementation. They comprised one Head of Foundation, one Arabic Language Program Coordinator, and six Arabic language teachers, for a total of eight interview participants. The quantitative component employed total sampling, involving all six Arabic language teachers as respondents (N = 6).

Participants were included if they were actively involved in Arabic language teaching or program management, had served at the institution for at least one year, and voluntarily agreed to participate in the study.

Evaluation Indicators and Data Collection

The evaluation indicators were derived from the decision-oriented functions of the Context and Input components of the CIPP framework. Context evaluation covered institutional alignment, evidence-based needs assessment, identification of program problems and priorities, and competency alignment and progression across educational units. Input evaluation covered teacher qualifications and professional readiness, instructional media and teaching materials, learning facilities and infrastructure, and curriculum and instructional documents.

Data was collected through semi-structured interviews, non-participant observations, a questionnaire, and document analysis. The relationship between data-collection techniques, sources, and evaluation purposes is presented in Table 1.

Table 1. Data Collection Techniques and Data Sources

No.	Technique	Data source	Number	Selection Technique	Evaluation Purpose
1.	Semi-structured interview	Head of Foundation, Arabic Language Program Coordinator, and Arabic language teachers	8 participants	Purposive sampling	Exploring institutional objectives, needs-assessment practices, competency alignment, program policies, teacher assignment, and resource

					conditions
2.	Non-participant observation	Classrooms, library, teaching materials, instructional media, audiovisual equipment, and learning-support facilities	Institutional facilities	Criterion-based observation	Assessing the availability, condition, diversity, and functionality of program resources
3.	Questionnaire	All Arabic teachers	6 respondents	Total sampling	Measuring the perceived adequacy of human resources, instructional resources, facilities, and curriculum support
4.	Documentation Study	Institutional, curriculum, teacher, program-planning, and facility documents	Relevant documents	Purposive document selection	Verifying institutional alignment, needs assessment, competency progression, teacher qualifications, curriculum readiness, and resource availability

Semi-structured interviews explored institutional objectives, needs-assessment practices, learning orientations across educational units, curriculum coordination, teacher assignment, resource availability, and program-development priorities. With the participants' consent, interviews were recorded and transcribed for analysis.

Non-participant observations were conducted using a structured observation sheet to examine the availability and condition of classrooms, library resources, teaching materials, instructional media, audiovisual equipment, and other learning-support facilities. The observations focused on program-resource readiness and were not used to evaluate teaching performance, classroom implementation, or student competence.

The questionnaire consisted of 24 items distributed equally across four Input dimensions: Teacher qualifications and professional readiness, six items; instructional media and teaching materials, six items; learning facilities and infrastructure, six items; and curriculum and institutional support, six items.

Responses were recorded on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores indicated a higher perceived level of input

adequacy. Document analysis was conducted to verify and complement evidence obtained from interviews, observations, and the questionnaire. The documents included the foundation's vision and mission statements, curricula, syllabi, lesson plans, teacher qualifications and assignment records, program-planning documents, and facility inventories.

Instrument Review, Reliability, and Trustworthiness

The questionnaire, interview guide, observation sheet, and document-analysis sheet were reviewed by one expert in educational evaluation. The expert was selected based on relevant academic qualifications and experience in program evaluation and educational instrument development. The review focused on the items' relevance to the Context and Input indicators, the clarity of wording, the representativeness of the evaluated dimensions, and their suitability for *pesantren*-based Arabic language education.

Items identified as unclear, overlapping, or insufficiently representative were revised based on the expert's recommendations before data collection. Because the review involved only one expert and no numerical Content Validity Index or Aiken's V coefficient was calculated, the procedure was reported as an expert review, rather than as statistical evidence of content validity.

The internal consistency of the 24-item questionnaire was examined using Cronbach's Alpha. The instrument produced an overall coefficient of $\alpha = 0.87$, indicating good internal consistency. This coefficient was used as evidence of questionnaire reliability, not of content validity.

The trustworthiness of the qualitative findings was strengthened through source and technique triangulation, member checking, and an audit trail. Interview evidence was compared with questionnaire results, observational findings, and institutional documents. Interview summaries were returned to participants to verify their accuracy, while the audit trail documented participant selection, data collection, transcription, coding, theme development, quantitative calculations, data integration, and the basis for each evaluative judgment. All participants were informed about the objectives and procedures of the study. Participation was voluntary, informed consent was obtained before data collection, and participant identities were protected by reporting institutional roles rather than personal names.

Data Analysis and Integration

Interview data were analyzed using thematic analysis through six stages: familiarization with the data, initial coding, theme development, theme review, theme definition, and report construction (Braun & Clarke, 2006). Deductive codes were derived from the Context and Input indicators, while inductive codes were used to capture relevant issues emerging from participants' responses.

Institutional documents were examined separately using qualitative content analysis. Relevant documentary evidence was extracted, categorized according to the Context and Input indicators, and compared with interview and observational findings.

Questionnaire data were analyzed using descriptive statistics. The achievement percentage for each Input dimension was calculated using the following formula:

$$\text{Achievement Percentage} = \frac{\text{Observed Score}}{\text{Ideal Score}} \times \mathbf{100}$$

The ideal score for each six-item dimension was calculated as follows:

$$\text{Ideal Score} = \mathbf{6}\ \text{respondents} \times \mathbf{6}\ \text{items} \times \mathbf{5}\ \text{maximum score} = \mathbf{180}$$

For descriptive interpretation, the percentage scale was divided into five equal intervals. The interval width was calculated by dividing the 0-100% range into five categories. The resulting interpretation criteria are presented in Table 2.

$$\text{Interval Width} = \frac{\mathbf{100 - 0}}{\mathbf{5}} = \mathbf{20}$$

Table 2.
Researcher-Developed Equal-Interval Interpretation Criteria

Achievement Percentage	Evaluative Category
81–100%	Very adequate
61–80%	Adequate
41–60%	Moderately adequate
21–40%	Inadequate
0–20%	Very inadequate

RESULT AND DISCUSSION

Result

Context Evaluation of Arabic Language Learning Programs

The Context evaluation examined the alignment of the Arabic language learning program with the foundation's vision and mission, the availability of an evidence-based needs assessment, and the alignment and progression of learning objectives across educational units. The program was considered contextually adequate when its objectives were institutionally relevant, supported by documented needs, and organized within a coherent competency framework.

Document analysis showed that Arabic language learning was explicitly positioned as an important component of the foundation's Islamic educational mission. Interview data from institutional leaders, the program coordinator, and teachers consistently indicated that Arabic was treated as a priority subject across the Qur'anic education, Madrasah Diniyah, and formal education units. No contradiction was identified between the general objectives of the Arabic language program and the foundation's vision and mission. Based on this evidence, institutional alignment was evaluated as adequate.

However, the document review did not identify a formal needs assessment document that contained baseline information on students' proficiency, expected competencies, stakeholder expectations, or the differentiated needs of each educational unit. Interview data indicated that the existing program had largely developed from established curriculum traditions and institutional experience. Although these traditions provided a basis for program continuity, they did not constitute a systematic empirical assessment of learning needs. This gap limited the institution's ability to demonstrate that program priorities were derived from students' current and expected competencies. Therefore, the needs-assessment dimension was evaluated as inadequate.

The evaluation also identified differences in learning orientations across educational units. Madrasah Diniyah emphasized reading and understanding classical Islamic texts, whereas formal education units addressed listening, speaking, reading, and writing competencies. These differences were not automatically judged as curricular misalignment because they reflected the distinct educational functions of the respective units. Nevertheless, the document analysis did not identify an explicit, foundation-wide framework that showed how the competencies developed by each unit were connected or progressively organized across educational levels. Thus, the identified gap concerned the absence of documented competency alignment and progression rather than the existence of differentiated objectives. This dimension was evaluated as partially adequate.

Overall, the Context component was evaluated as moderately adequate. The program possessed strong institutional legitimacy and clear alignment with the foundation's educational vision. Still, the absence of a formal needs assessment and an explicit cross-level competency framework limited its contextual readiness.

Input Evaluation of the Arabic Language Learning Program

The Input evaluation examined the adequacy of teacher qualifications and professional readiness; instructional media and teaching materials; learning facilities and infrastructure; and curriculum and instructional documents. Quantitative questionnaire results were integrated with evidence from interviews, observations, and document analysis. The percentage scores were interpreted using the evaluation categories established in the Methods section.

Table 3.
Evaluation Results for the Input Components

No.	Input Dimension	Achievement	Category
1.	Teacher qualifications and professional readiness	45,0%	Moderately adequate
2.	Instructional media and teaching materials	58%	Moderately adequate
3.	Learning facilities and infrastructure	62,5%	Adequate
4.	Curriculum and instructional documents	71%	Adequate
Average		**59,1%**	**Moderately adequate**

The teacher qualifications and professional readiness dimension achieved a score of 45.0%. Document verification showed that only a limited proportion of the teachers possessed academic backgrounds directly related to Arabic Language Education or Arabic Literature. Interview evidence indicated that teaching assignments were also based on seniority, institutional experience, alum status, and the ability to read Islamic texts. The integration of questionnaire, interview, and documentary data revealed a gap between the specialized competencies expected for Arabic language instruction and the formal qualifications held by the teaching staff. This dimension was evaluated as moderately adequate but requiring priority improvement. This judgment does not imply that teachers with non-aligned academic backgrounds were ineffective because classroom teaching performance was not evaluated in this study.

Instructional media and teaching materials achieved 58.0%. Observations showed that printed textbooks, blackboards, and conventional materials remained the predominant resources. Digital applications, audiovisual resources, presentations, and interactive teaching materials were available only to a limited extent. The observational evidence supported the questionnaire findings, indicating convergence between the quantitative and qualitative datasets. The available materials supported basic teaching activities but were not sufficiently diverse to facilitate the intended range of language competencies, particularly listening and speaking. This dimension was therefore evaluated as moderately adequate.

Learning facilities and infrastructure achieved 62.5%. Observations showed that classrooms were generally functional and that the library contained Arabic-language resources. However, most library materials consisted of classical texts, while contemporary language-learning resources and audiovisual equipment remained limited. The institution did not have a functioning language laboratory. However, this absence was not treated as evidence that listening and speaking competencies could not be developed through other resources and communicative environments. The evaluative gap concerned the limited variety of facilities available to support sustained oral and aural language practice. This dimension was evaluated as adequate but in need of improvement.

Curriculum and instructional documents achieved the highest score at 71.0%. Document analysis confirmed the availability of curricula, syllabi, and lesson plans across the educational units. These documents generally corresponded with applicable administrative and curricular requirements. However, their availability was not accompanied by equivalent readiness in teacher qualifications, instructional media, and supporting facilities. This dimension was adequate, but we need stronger alignment with the needs assessment findings and cross-level competency progression.

Discussion

Context Evaluation of Arabic Language Learning Programs

The Context evaluation revealed a contrast between strong institutional commitment and limited evidence-based program planning. Arabic language learning had a legitimate and strategic position within the educational vision of the Tarbiyatut

Tholabah Foundation. However, this legitimacy had not yet been fully translated into a systematic mechanism for identifying learning needs, setting priorities, and organizing competency progression.

Within the CIPP framework, contextual adequacy is not determined solely by the formal compatibility between program objectives and institutional vision. Context evaluation also requires identifying beneficiary needs, institutional problems, available opportunities, and program priorities (Stufflebeam & Zhang, 2017). Therefore, the absence of a documented needs assessment constitutes a substantive limitation, even though the program's general objectives remain relevant to the foundation's Islamic educational mission.

This finding is comparable with studies reporting that Arabic language curricula in *pesantren* are commonly shaped by normative objectives and established educational traditions (Fathoni et al., 2021; Samsuddin & Siregar, 2024). However, the present study extends previous work by locating a critical challenge at the planning stage rather than solely within curriculum implementation or learning outcomes. Program limitations may emerge before instruction begins when decision-makers lack documented evidence of student proficiency, stakeholder expectations, and unit-specific learning priorities.

The absence of a formal needs assessment does not prove that the existing objectives are inappropriate. Rather, it limits the institution's ability to demonstrate that its objectives correspond to actual and evolving learner needs. A structured needs assessment would enable the foundation to differentiate among religious, academic, and communicative language requirements and establish the relative priority of each competency at different educational levels.

Differences in learning orientation across educational units also require cautious interpretation. The Madrasah Diniyah's emphasis on reading classical Islamic texts is consistent with its religious and scholarly mission. Meanwhile, the emphasis in formal educational units on listening, speaking, reading, and writing aligns with broader curricular requirements. These differences do not automatically indicate misalignment.

The evaluative concern is the absence of an explicit institutional framework that connects the differentiated objectives. Without a cross-level competency roadmap, there is a potential risk of overlap, discontinuity, or insufficient progression across educational levels. Hence, the issue is not the differentiation itself but whether it is deliberately organized within a coherent institutional pathway.

This interpretation extends previous curriculum studies by distinguishing appropriate differentiation from uncoordinated fragmentation. A responsive Arabic language curriculum does not necessarily impose identical objectives on all educational units. Instead, it should preserve unit-specific purposes while defining common foundational competencies, progression expectations, and points of connection across educational levels (Abdurrahman, 2025; Adlya et al., 2025).

The qualitative data converged on this conclusion. Documents showed a clear institutional commitment; interviews confirmed Arabic as a priority subject; and both sources indicated the absence of a formal needs assessment and a cross-level competency framework. The principal Context limitation was therefore not weak institutional

commitment but the limited translation of that commitment into diagnostic and coordinated planning.

Input Evaluation of the Arabic Language Learning Program

The Input evaluation demonstrated uneven readiness across human resources, media, facilities, and curriculum documentation. The overall score of 59.1% fell within the moderately adequate category. This judgment reflects the presence of basic program resources, but several dimensions need substantial strengthening.

Teacher qualification alignment represented the lowest-scoring dimension. This finding should not be interpreted as evidence that teachers with non-aligned academic backgrounds necessarily perform ineffectively. The study did not directly evaluate classroom performance or compare teaching quality according to academic background. Moreover, teaching experience, mastery of classical Islamic texts, and familiarity with *pesantren* culture may constitute valuable competencies that are not fully represented by formal qualifications.

Nevertheless, specialized academic preparation generally provides exposure to language pedagogy, curriculum development, assessment, and contemporary instructional approaches. Previous research has emphasized that Arabic language teaching requires a combination of language knowledge, pedagogical competence, and professional confidence (Ahmad et al., 2024; Tauhid, 2026). The current findings extend this concern from individual competence to institutional staffing policy. The issue is not merely whether teachers hold relevant degrees, but whether recruitment, assignment, and professional development are systematically aligned with program objectives.

The mixed-methods evidence reinforced this interpretation. Questionnaire responses indicated limited human-resource readiness, teacher-profile documents showed low alignment with qualifications, and interviews revealed that teaching assignments were also shaped by institutional experience, alum status, and classical-text competence. These sources converged in demonstrating that traditional *pesantren* staffing strengths had not yet been fully integrated with specialized professional criteria.

Instructional media and teaching materials were also moderately adequate. The continued dominance of printed materials and conventional classroom media indicates that technology integration remains limited. Previous studies have shown that digital and interactive media can increase exposure to authentic language, expand learning activities, and support learner engagement (Abid & Sain, 2025; Faiqoh et al., 2025; Zaenuri et al., 2025).

The implication, however, is not that conventional materials should be replaced entirely. Classical texts and printed resources remain central to particular *pesantren* learning objectives. Digital and audiovisual resources should complement these materials by supporting competencies that are less easily developed through text-based instruction, particularly listening, pronunciation, and spoken interaction.

The questionnaire identified the perceived level of media adequacy, while observation clarified that the limitation concerned the range and functionality of the available resources. This represents complementarity between the quantitative and

qualitative datasets. The main issue was not the complete absence of materials but their limited diversity in relation to the intended language competencies.

Facilities and infrastructure achieved a higher score, although limitations remained. The availability of classrooms and library resources provided a foundation for learning. However, audiovisual facilities and specialized language-learning resources were limited.

A language laboratory should not be treated as the only mechanism for developing listening and speaking skills. These competencies can also be supported through classroom interaction, peer communication, mobile devices, audio recordings, language clubs, and extracurricular language environments. Thus, the evaluative issue concerns whether learners have sufficient opportunities to receive spoken-language input and engage in meaningful communication rather than whether the institution possesses a particular physical facility.

Curriculum and instructional documents represented the strongest Input dimension. This finding differs from studies in which formal curriculum incompleteness constitutes the principal weakness. At the Tarbiyatut Tholabah Foundation, curricula, syllabi, and lesson plans were generally available, but the resources required to operationalize them remained less developed.

This contrast illustrates the distinction between administrative and operational readiness. Formal documents may comply with curricular requirements, but effective program readiness also depends on teacher expertise, instructional resources, and facilities. The quality of curriculum documents does not automatically guarantee implementation capacity when the supporting inputs remain uneven (Samsuddin & Siregar, 2024).

Overall, the Input findings indicate that the program is not without resources. Rather, its readiness is uneven. To strengthen the program, we need targeted interventions rather than a blanket conclusion that all input dimensions are inadequate.

Interpretive Linkages between Context and Input Evaluation Findings

Although this study did not test causal or statistical relationships between the Context and Input components, the integrated findings indicate several plausible linkages between contextual diagnosis and resource readiness. These linkages are interpreted as evaluative patterns rather than as evidence of direct causation.

The absence of a formal, evidence-based needs assessment appears to be associated with the limited availability of explicit benchmarks for program-resource planning. Without clear information on students' current proficiency, expected competencies, stakeholder priorities, and the distinctive objectives of each educational unit, institutional leaders may struggle to determine the types of teacher expertise, instructional media, teaching materials, and facilities the program needs. Consequently, resource decisions may be shaped more strongly by established institutional practices and immediate availability than by documented competency requirements.

A similar linkage was identified between differentiated learning orientations and input requirements. The emphasis on classical-text reading in Madrasah Diniyah and the

broader focus on listening, speaking, reading, and writing in formal educational units represent different but potentially legitimate instructional orientations. However, these orientations also require different combinations of teacher competence, instructional resources, and learning facilities. In the absence of an explicit cross-level competency framework, the foundation lacks a documented basis for determining which inputs should be shared across educational units and which should be differentiated according to their respective purposes.

The integrated mixed-methods findings support this interpretation. Document analysis confirmed the availability of curriculum and instructional documents but did not identify a formal needs assessment or a foundation-wide competency-progression framework. Questionnaire data indicated uneven input readiness, particularly in teacher qualification alignment and instructional-media provision. Observation clarified the limited diversity of audiovisual and technology-supported resources. At the same time, interview data explained that teacher assignment and resource decisions were influenced by institutional experience, alum status, and established *pesantren* traditions. The convergence and complementarity of these findings suggest a potential linkage between limited contextual diagnosis and uneven resource planning.

Teacher qualification alignment should nevertheless be interpreted cautiously. The findings do not demonstrate that teachers with non-aligned academic backgrounds employ ineffective or traditional teaching methods, because classroom practices were not systematically evaluated. Experience in *pesantren* education, mastery of classical Islamic texts, and familiarity with institutional culture may constitute important forms of professional expertise. However, the gap between academic specialization and teaching responsibilities indicates a need for more systematic professional-development planning and for future research examining how different forms of teacher expertise are applied in classroom instruction.

Similarly, limited audiovisual resources and the absence of a functioning language laboratory should not be interpreted as direct evidence of inadequate student language competence. Listening and speaking skills can be supported through various resources and learning environments, including classroom interaction, audio recordings, mobile devices, peer communication, language clubs, and structured extracurricular activities. The current evaluation shows that the range of institutional resources available to support these activities remains limited.

These findings illustrate the decision-oriented logic of the CIPP framework, in which Context evaluation identifies program needs and priorities. In contrast, Input evaluation examines the resources and strategies available to address them. The present study does not establish that weaknesses in Context caused limitations in Input. Instead, it suggests that stronger contextual diagnosis could provide a clearer foundation for teacher recruitment, professional development, curriculum coordination, media development, and infrastructure planning.

Therefore, the principal implication arising from the relationship between the Context and Input findings concerns institutional planning rather than program effectiveness or student achievement. A formal needs assessment and a clear cross-level

competency framework can help the foundation develop more transparent criteria for teacher assignment, resource allocation, instructional-media selection, and facility development. The effects of these improvements on teaching practices and student learning outcomes should be examined separately through future evaluations of the Process and Product components.

CONCLUSION

The evaluation indicates that the Arabic language learning program at the Tarbiyatut Tholabah Foundation has strong institutional legitimacy but incomplete contextual and input readiness. In the Context component, the program aligns with the foundation's educational vision. Still, it lacks a formal needs assessment or an explicit framework to align and advance Arabic competencies across educational levels. In the Input component, the overall achievement score was 59.1%, categorized as moderately adequate. Curriculum documents were relatively well prepared, while teacher qualification alignment, instructional media, and supporting facilities remained priority areas for improvement.

This study provides a diagnostic basis to strengthen the program by conducting an evidence-based needs assessment, mapping cross-level competencies, aligning teacher expertise with teaching responsibilities, supporting continuous professional development, and developing targeted instructional resources. Because the evaluation was limited to Context and Input, the conclusions concern program needs and institutional readiness and should not be interpreted as evidence of effects on instructional processes or student learning outcomes. Future studies should evaluate the Process and Product components after the recommended improvements have been implemented.

REFERENCES

Abdilah, A. J. (2025). Evaluasi Kurikulum Bahasa Arab dengan Model CIPP (Context, Input, Process, and Product). *Risenologi*, *10*(1), 52–68. https://doi.org/10.21009/risenologi.101.06

Abdurrahman, A. (2025). Evaluasi Dan Pengembangan Kurikulum Pendidikan Bahasa Arab: Analisis Berbasis Model CIPP. *Jurnal Manajemen Pendidikan*, *10*(4), 2974–2986. https://doi.org/10.34125/jmp.v10i4.1381

Abid, H. A., & Sain, Z. H. (2025). The Transformation of Arabic Language Learning in the Digital Era: A Critical Review of Technological Innovation, Linguistic Complexity, and the Pedagogical Imperatives of TPACK (2020–2025). *JETech: Journal of Education and Technology*, *1*(3), 97–106. https://doi.org/10.65678/jetech.v1i3.231

Adlya, F. I., Salsabila, A., & Baity, A. (2025). Implementasi Model Evaluasi CIPP untuk Mewujudkan Pembelajaran Bahasa Arab Holistik dalam Kurikulum Merdeka. *EL-FUSHA: Jurnal Bahasa Arab dan Pendidikan*, *6*(2), 206–218. https://doi.org/10.33752/el-fusha.v6i2.10162

Ahmad, F. S., Yasin, A., Sofian, A. A., Khalifah, N. F., & Rahmawati, N. (2024). Curriculum Enhancement and Language Assessment in *Pesantren*: Examining Arabic Oral and Written Tests at Pondok Modern Darussalam Gontor Putri Campus 1. *AFKARINA:*

Jurnal Pendidikan Agama Islam, *9*(1), 71–82. https://doi.org/10.33650/afkarina.v9i1.15528

Akhyar, K. N., Suhardi, I., & Patahuddin. (2025). Evaluation of The Arabic Language Learning Program based on The Merdeka Curriculum at a Madrasah Tsanawiyah in Makassar. *Lentera Pendidikan : Jurnal Ilmu Tarbiyah Dan Keguruan*, *28*(1), 40–67. https://doi.org/10.24252/lp.2025v28n1i3

Arofah, E. F. (2021). Evaluasi Kurikulum Pendidikan. *Jurnal Tawadhu*, *5*(2), 218–229. https://doi.org/10.52802/twd.v5i2.236

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and Conducting Mixed Methods Research*. SAGE Publications.

Faiqoh, P. K., Gadoum, H., Baity, A. N., & Azza, F. H. (2025). Digital Media and Technology in Arabic Language Learning. *Journal of Arabic Language Teaching*, *5*(1), 121–132. https://doi.org/10.35719/arkhas.v5i1.2263

Fathoni, W., Alim, K., N, Y., & Latif, M. (2021). Evaluation of Islamic Education Curriculum. *Al-Mubin: Islamic Scientific Journal*, *4*(1), 108–116. https://doi.org/10.51192/almubin.v8i2.2036

Fitri, N., Kiram, S., Suwadi, Priyanggani, W., & Jabil, C. L. (2026). Evaluating an Arabic–English Bilingual Program in an Islamic Boarding School Through the Cipp Model: Strengthening Students' Language Competence. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, *11*(02), 228–253. https://doi.org/10.23969/jp.v11i02.52692

Hadisi, L., & Muna, W. (2025). Analisis Manajemen Pembelajaran Nahwu-Sharaf di Pondok *Pesantren* Al Jannah (Studi pada Pembelajaran Bahasa Arab di Lingkungan *Pesantren*). *Banggona Metulura*, *1*(2), 44–54.

Muhajir, M. (2021). Ībistīmūlūjiyya Al-Manāhij al-Dirāsiyyah fī *Pesantren*: Ru'yah Tajdīdiyyah li Ta'limi al-Lugah al-'Arabiyyah / Epistemology Curriculum in *Pesantren* Indonesia: A New Vision for Arabic Learning. *Al Mahāra: Jurnal Pendidikan Bahasa Arab*, *7*(2), 269–284. https://doi.org/10.14421/almahara.2021.072-06

Samsuddin, S., & Siregar, I. S. (2024). Optimizing Islamic Religious Education: A Case Study on Curriculum Management in Planning, Implementation, and Evaluation. *AL-ISHLAH: Jurnal Pendidikan*, *16*(4), 5150–5162. https://doi.org/10.35445/alishlah.v16i4.6389

Sapura, N. L., Amalia, L., Shofah, A., & Nugraha, R. M. (2025). Evaluasi Implementasi Kurikulum Bahasa Arab Di Pondok *Pesantren* Salafiyah Dan Modern. *MANAJERIAL : Jurnal Inovasi Manajemen Dan Supervisi Pendidikan*, *5*(3), 718–727. https://doi.org/10.51878/manajerial.v5i3.6944

Stufflebeam, D. L., & Zhang, G. (2017). *The CIPP Evaluation Model: How to Evaluate for Improvement and Accountability*. Guilford Publications.

Tauhid, T. (2026). Transforming Arabic Learning in Indonesian *Pesantren*: Policy, Teacher Competence, and Reform Challenges. *Tanwir Arabiyyah: Arabic As Foreign Language Journal*, *6*(1), 1–19. https://doi.org/10.31869/aflj.v6i1.8039

Zaenuri, M., Yunus, M., Sya'bani, M. Z., & Ahmad, Z. A. (2025). Exploring Students' Preferences and Satisfaction in Using Digital Media for Arabic Language Learning in Islamic Higher Education. *ATHLA : Journal of Arabic Teaching, Linguistic and Literature*, *6*(1), 103–127. https://doi.org/10.22515/athla.v6i1.11998