



Curriculum Alignment, Implementation Fidelity, and Outcome Evidence in a Pesantren-Based Intensive English Course: A Discrepancy Evaluation of the Madani Course Program

Muh Junaidi¹, Ni Nyoman Padmadewi¹, I G A Lokita Purnamika Utami¹, Putu Suarcaya¹
¹ Universitas Pendidikan Ganesha, Bali, Indonesia

Article Info

Article history:

Received Jun 12, 2026
Revised Jun 30, 2026
Accepted Jul 23, 2026
OnlineFirst Jul 31, 2026

Keywords:

Curriculum Alignment
Discrepancy Evaluation Model
Implementation Fidelity
Intensive English Course
Pesantren

ABSTRACT

Purpose of the study: The Intensive English programs in pesantren often provide rich exposure and strong motivational support, yet their curriculum, implementation, assessment, and documented outcomes may not be systematically aligned. This study evaluated curriculum alignment, implementation fidelity, outcome evidence, and improvement priorities in the four-week Madani Course Program using the five stages of the Discrepancy Evaluation Model (DEM).

Methodology: A qualitative evaluative case study involved one program director, one coordinator, four tutors, and twenty-four students. Data were generated through semi-structured interviews, six classroom observations, and analysis of schedules, materials, attendance records, tutor notes, assessment records, and program reports. The evidence was thematically coded, triangulated in student- and program-level evidence matrices, compared with explicit standards, and classified as minor, moderate, or major discrepancies.

Main Findings: Triangulated evidence indicated positive perceived and observed outcomes in confidence, participation, routine vocabulary use, self-introduction, and learning discipline. However, moderate discrepancies affected measurable outcomes, curriculum sequencing, tutor preparation, structured modules, target-language use, feedback, monitoring, and reporting. A major product-stage discrepancy remained because comparable diagnostic and final speaking assessments were unavailable; therefore, firm claims about linguistic improvement could not be made.

Novelty/Originality of this study: The study contextualizes DEM by integrating communicative pedagogy, formative assessment, and pesantren educational culture into a transparent program-alignment framework. The findings imply that similar residential programs should connect measurable weekly outcomes, tutor preparation, staged language-use policies, comparable speaking tasks, portfolios, and evidence-based reporting.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license
© 2026 by the author(s)



Corresponding Author:

Muh Junaidi,
Universitas Pendidikan Ganesha, Jalan Udayana No.11. Singaraja, Bali, 81116, Indonesia
Email: muhjunaidi13@gmail.com

1. INTRODUCTION

Pesantren have continued to adapt to changing educational, social, and technological demands while preserving their religious, moral, residential, and communal foundations. In contemporary Indonesia, many

pesantren combine religious study with formal schooling, digital literacy, vocational preparation, and foreign-language education. English is increasingly positioned as a resource for academic mobility, access to global knowledge, institutional visibility, and intercultural communication. Recent studies show that this expansion is accompanied by persistent questions about teacher readiness, language policy, learning materials, student motivation, and the capacity of pesantren to reconcile innovation with their distinctive educational culture [1]-[8].

The residential structure of a pesantren can offer conditions that are difficult to reproduce in ordinary English as a foreign language (EFL) classrooms. Students live together, follow common schedules, interact repeatedly with peers, and can use English beyond formal lessons. Research has associated these conditions with increased exposure, confidence, vocabulary practice, participation, and language habituation [2], [9]-[15]. However, the existence of an English environment does not by itself demonstrate program quality. Intensive schedules may generate many activities without a coherent progression of learning outcomes; enthusiastic tutors may use different instructional approaches; and confidence may improve without reliable evidence of development in fluency, pronunciation, vocabulary, grammar, or interaction. The central evaluation problem is therefore not merely whether activities occur, but whether program intentions, resources, implementation, outcomes, and reported benefits are aligned.

The Madani Course Program at Pondok Pesantren Nurul Haramain Lombok represents this type of intensive, residential English initiative. The program seeks to strengthen students' willingness and ability to use English through scheduled instruction, peer practice, language routines, and pesantren discipline. Previous work at Nurul Haramain has documented motivational factors that support participation in English programs [12]. Nevertheless, a motivational description cannot establish whether the course has measurable outcomes, a shared curriculum sequence, adequately prepared tutors, consistent feedback, or sufficient evidence for accountability. A whole-program evaluation is required to identify where expected standards and actual performance converge or diverge.

The Discrepancy Evaluation Model (DEM) provides an appropriate logic for this purpose. DEM defines evaluation as a systematic comparison between standards and performance and uses the identified discrepancy as diagnostic information for decision-making [16], [17]. Its stages address program design, installation, process, product, and cost-benefit considerations. Design concerns the clarity of goals, learning outcomes, curriculum structure, and assessment planning. Installation examines whether tutors, materials, facilities, schedules, rules, and monitoring systems are ready. Process considers the fidelity and quality of implementation. Product concerns evidence of intended outcomes. Cost-benefit evaluation considers whether documented benefits justify the resources, time, and institutional investment. Although DEM is well established, its application requires explicit, context-sensitive standards rather than mechanical judgements [18]-[20].

For an intensive English course, the standards must also reflect what meaningful language development entails. Communicative competence extends beyond memorized vocabulary and grammatical knowledge to the ability to use language appropriately and effectively in social interaction [21], [22]. Communicative tasks require clear goals, relevant input, modeling, rehearsal, interaction, scaffolding, and opportunities for increasingly independent performance [23]-[28]. Program outcomes should therefore distinguish participation and confidence from demonstrable speaking performance. A supportive classroom climate is important because anxiety can restrict achievement and willingness to communicate, whereas positive teacher and peer environments can promote engagement [29]-[35]. At the same time, confidence should be accompanied by assessment evidence. Formative assessment and feedback are most useful when they clarify the expected standard, indicate the learner's current performance, and guide the next improvement step [36]-[40].

The recent literature on English in pesantren is growing, but it remains fragmented. Studies have examined general development and challenges [7], instructional strategies [3], teacher difficulties [6], motivational participation [12], environmental and institutional support [15], linguistic schoolscape [9], context-responsive pedagogy [10], learning materials [4], language policy [1], microlearning [41], and residential speaking environments [11]. Evaluative studies have mainly used the Context, Input, Process, and Product model to describe integrated or bilingual programs [42], [43], while other program evaluations have focused on broader pesantren education [44]. These studies are valuable, yet four gaps remain. First, the Madani Course has not been evaluated as an interconnected program system. Second, few studies use DEM to compare explicit standards and actual performance across all five stages. Third, previous research often reports perceptions and activities without constructing a transparent discrepancy map that links evidence to improvement priorities. Fourth, the literature has not sufficiently integrated program evaluation standards with communicative pedagogy, formative assessment, and pesantren culture.

The novelty of this study lies in addressing these gaps through an integrated program-alignment framework. DEM supplies the evaluative architecture; communicative language learning and formative assessment supply pedagogical and outcome standards; and pesantren culture supplies the contextual lens. This integration enables the evaluation to examine five alignments: institutional vision with measurable learning outcomes; program resources with expected readiness; classroom activities with communicative goals; perceived

progress with documented learning evidence; and reported benefits with stakeholder accountability. Rather than labeling the program simply successful or unsuccessful, the framework identifies which components should be maintained, revised, standardized, or documented.

Accordingly, this study aimed to evaluate curriculum alignment, implementation fidelity, outcome evidence, and improvement priorities in the Madani Course Program. It addressed three research questions: (1) How does the Madani Course Program align with the expected standards of a pesantren-based intensive English course across the DEM stages of design, installation, process, product, and cost-benefit? (2) What major discrepancies are evident across these stages? (3) What improvement priorities can strengthen program quality, accountability, and sustainability?

2. RESEARCH METHOD

This study employed a qualitative evaluative case-study design guided by DEM. The case was bounded by one four-week implementation of the Madani Course Program, its participants, instructional activities, management arrangements, and available program documents. The evaluative orientation distinguished the study from a purely descriptive case study: evidence was interpreted by comparing actual conditions with contextually appropriate standards and by converting identified discrepancies into improvement priorities [16], [17], [20]. The qualitative approach was selected because the evaluation required detailed accounts of program intentions, stakeholder experiences, classroom processes, institutional conditions, and the meaning of available outcome evidence.

The five DEM stages were operationalized before data analysis. Design covered program aims, measurable learning outcomes, curriculum sequence, teaching approach, and assessment planning. Installation covered tutor readiness, materials, facilities, schedules, language rules, and monitoring tools. Process covered instructional implementation, English use, student participation, scaffolding, feedback, and monitoring. Product covered perceived and documented changes in confidence, participation, vocabulary use, speaking performance, and discipline. Cost-benefit covered the program's value in relation to time, staffing, facilities, accountability, and sustainability.

The evaluation was conducted at Pondok Pesantren Nurul Haramain Lombok, West Nusa Tenggara, during a four-week Madani Course cycle. The course used a residential arrangement in which students attended scheduled English sessions, practiced with peers, and participated in routines supervised by tutors and program management. The students were predominantly at beginner and lower-intermediate levels according to tutor judgement and classroom observation.

Purposive sampling was used because participants needed direct knowledge of program planning, management, teaching, or participation [45]. The sample included the complete program-management team involved in the cycle (one director and one coordinator), all four English tutors, and twenty-four students. Participant codes protected confidentiality: PD for the program director, PC for the coordinator, T1-T4 for tutors, and S1-S24 for students.

Table 1. Participant Profile

Participant group	Number	Role in the evaluation	Code
Program director	1	Program vision, institutional policy, expected outcomes, and intended benefits	PD
Program coordinator	1	Management, schedule, resources, monitoring, documentation, and reporting	PC
English tutors	4	Preparation, materials, teaching strategies, feedback, assessment, and implementation challenges	T1-T4
Student participants	24	Learning experience, participation, confidence, difficulties, perceived progress, and program value	S1-S24

Conventional statistical power was not applicable because this was a qualitative, single-case evaluation that did not test group differences or estimate population effects. Sample adequacy was instead judged through information power: the study had a narrow aim, a highly specific participant group, an established evaluation framework, multiple stakeholder perspectives, repeated observations, and document evidence [46]-[48]. The inclusion of all managers and tutors, together with twenty-four students, six observations, and program documents, provided sufficient case-level evidence for triangulated analysis. The findings are intended for analytical transferability to comparable programs, not statistical generalization.

Data were generated through semi-structured interviews, classroom observations, and document analysis during the four-week course. Interviews were conducted in Indonesian so that participants could express detailed views. The director and coordinator were interviewed for approximately 45-60 minutes each; tutors for 30-45 minutes each; and students for approximately 15-25 minutes each. With consent, interviews were audio-

recorded and transcribed in Indonesian. Selected quotations were translated into English for reporting. Six teaching sessions, each lasting approximately 60-90 minutes, were observed. The sessions included vocabulary practice, guided conversation, pronunciation work, role play, short performance, and review. Structured field notes recorded interaction patterns, target-language use, task organization, student participation, scaffolding, feedback, time management, and monitoring. Documents included schedules, attendance lists, teaching materials, tutor notes, available assessment records, and program reports. Table 2 aligns the data-collection instruments with the DEM stages and the evidence sought.

Table 2. Data Collection Instrument Grid

DEM stage	Key indicators	Primary data source	Instrument/technique	Evidence sought
Design	Goals; measurable outcomes; curriculum sequence; teaching approach; assessment plan	Director; coordinator; tutors; program documents	Management/tutor interview guides; document review form	Consistency between program vision, written plans, weekly targets, tasks, and assessment expectations
Installation	Tutor readiness; modules; facilities; schedule; rules; monitoring tools	Coordinator; tutors; schedules; materials; facilities	Coordinator/tutor interview guides; document review; observation checklist	Availability and adequacy of human, material, organizational, and monitoring resources
Process	Communicative activity; English use; participation; scaffolding; feedback; monitoring	Tutors; students; six classroom sessions; tutor notes	Tutor/student interview guides; structured observation field notes	Fidelity, consistency, and quality of classroom implementation across tutors and sessions
Product	Confidence; participation; vocabulary use; self-introduction; spontaneous response; pronunciation; discipline	Students; tutors; observations; attendance; speaking tasks; notes	Student/tutor interviews; outcome evidence sheet; document review	Convergence between claimed outcomes and evidence from at least two data sources
Cost-benefit	Perceived value; use of resources; reporting; accountability; sustainability	Director; coordinator; tutors; available reports	Management interview guide; document review form	Whether benefits are documented sufficiently to justify time, staffing, facilities, and continuation

Expected standards were derived from DEM, recent program-evaluation studies, communicative language pedagogy, formative-assessment principles, and the characteristics of residential pesantren learning [18], [19], [25]-[27], [36], [39], [40], [42], [43]. The standards were also checked against available program documents and stakeholder expectations. The program coordinator and one language-education colleague reviewed the standards for theoretical relevance, contextual suitability, and clarity. Their feedback was used to reduce vague judgement and to ensure that the criteria could be applied to a beginner-oriented intensive English course.

Discrepancies were classified as minor, moderate, or major. A minor discrepancy indicated a limited gap that did not substantially compromise implementation. A moderate discrepancy affected consistency, instructional quality, resource readiness, or monitoring. A major discrepancy directly weakened outcome evidence, accountability, or sustainability. Classification required explicit comparison between the expected standard and triangulated evidence; a positive perception alone was not treated as proof that a standard had been achieved.

Table 3. DEM Standards and Discrepancy Decision Rules

DEM stage	Expected standard	Evidence rule	Discrepancy focus
Design	Clear goals, measurable weekly outcomes, curriculum map,	Written or consistently articulated standards	Clarity and operationalization of

DEM stage	Expected standard	Evidence rule	Discrepancy focus
	teaching approach, and assessment plan	confirmed across management, tutors, and documents	program intentions
Installation	Prepared tutors, structured modules, facilities, schedule, rules, rubrics, and monitoring tools	Resource availability plus evidence that resources were ready for use before or during implementation	Readiness and implementation support
Process	Active, communicative, scaffolded, student-centered, feedback-rich, and monitored learning	Observed practice supported by tutor/student accounts and available monitoring records	Consistency, target-language use, participation, feedback, and monitoring
Product	Documented affective, behavioral, and linguistic outcomes	Clear evidence supported by at least two sources; limited evidence supported by one source or inconsistent evidence	Difference between claimed outcomes and available evidence
Cost-benefit	Documented benefits that justify time, staffing, facilities, and institutional investment	Stakeholder judgements supported by reports, records, portfolios, or comparable evidence	Reporting, accountability, and sustainability

Analysis combined thematic analysis with a DEM standards-performance matrix [50]-[53]. First, interview transcripts, observation notes, and documents were read repeatedly and organized by source and participant group. Second, initial codes were assigned to meaningful segments; examples included unclear outcomes, uneven tutor preparation, incomplete modules, inconsistent English use, supportive climate, limited formative feedback, perceived confidence, and weak outcome documentation. Third, related codes were grouped into candidate themes. Fourth, themes and evidence were mapped to the five DEM stages. Fifth, the expected standard and actual condition for each indicator were compared in a matrix. Sixth, discrepancies were classified according to the decision rules in Table 3. Seventh, improvement priorities were generated by considering the severity of each discrepancy, its influence on subsequent DEM stages, and its practical feasibility.

Triangulation was embedded in the analysis. Interview claims were checked against observations and documents, while management, tutor, and student perspectives were compared. For the product stage, 'clear evidence' required support from at least two sources, such as an interview and observation, a tutor note and performance task, or attendance and observation. 'Limited evidence' referred to a claim supported by only one source or appearing inconsistently across the corpus. This rule was used to produce the descriptive outcome counts reported in Table 4. No inferential statistical analysis was performed. The numerical counts in the findings are descriptive summaries created after qualitative classification; they are not standardized pre-test-post-test scores, independent quantitative measurements, or estimates of population parameters. Consequently, p-values, confidence intervals, effect sizes, and conventional statistical power calculations would be inappropriate. The counts are presented only to make the distribution of triangulated case evidence more transparent.

Credibility was strengthened through source and method triangulation, member checking of selected interview summaries, and peer debriefing. The researcher maintained an audit trail that linked standards, codes, evidence extracts, discrepancy judgements, and recommendations. Peer discussion with a language-education colleague examined whether classifications were reasonable and sufficiently supported. These procedures addressed the credibility, dependability, confirmability, and transparency expected in qualitative analysis [49], [52], [53]. Institutional permission was obtained before data collection. Participants received information about the study purpose, voluntary participation, audio recording, confidentiality, and academic use of the data. Identifying information was replaced with codes, and student identities were not reported individually.

3. RESULTS AND DISCUSSION

The results are organized by the five DEM stages. The stage-specific analysis answers the first research question; the discrepancy synthesis and matrix answer the second; and the ranked improvement priorities answer the third. Each result reports the expected standard, triangulated actual condition, discrepancy level, and practical consequence.

Design Stage: Contextually Strong Vision but Incomplete Operationalization

A matrix documentary program explaining the context was thus not rationally framed. According to both management and tutors, the course came about in response to students who were unable to have other real-

time opportunities for English use outside of regular classes. The outputs expectations featured English development as well as confidence, courage, discipline and learning habits. This vision was in line with the practice of pesantren education, which consists of a combination of academic studies, character building, communal activities, and repetitive exercises.

The main area of disparity was with how that high level vision would trickle down into measurable results. After each programme, program documents had made a broad expression such as improving English and confidence level and grabbing English environment. They did not regular say weekly language functions or performance objectives like performing a two-minute self-summary, addressing and answering plain inquiries of daily situations for illustration outlining day-to-day life practices taking part in an assisted discussion providing a quick readied talk. Consequently, the tutors had different interpretations of the aims of the program. In one class students memorized vocabulary, in another they played games, and in another the short speech was put first. The range of activities was perhaps beneficial, but the activities were not systematically linked to a common outcome map.

A tutor explained, 'We know that the students must be more confident speaking English, but we decide the daily focus depending on the situation. There is no fixed weekly target' (T2). This flexibility allowed tutors to respond to learners, yet it also reduced curriculum coherence. The materials included vocabulary lists, everyday conversations, pronunciation practice, role play, language games, and speech activities, but the sequence from controlled practice to guided interaction and freer performance was not explicit. A student commented, 'I like speaking activities, but sometimes I am confused because yesterday we learned vocabulary and today we have to make a speech. I need more examples first' (S8). The design stage was therefore rated as moderately discrepant. Contextual relevance and institutional commitment were strong, but measurable outcomes, curriculum mapping, task sequencing, assessment planning, and written standards were incomplete. The first design priority is an outcome-based curriculum that specifies weekly language functions, vocabulary themes, task progression, and assessment evidence.

Installation Stage: Strong Residential Resources but Uneven Pedagogical Readiness

Installation strengths of this program included the following areas. Residential setting facilitated regular attendance, prolonged contact time with peers to practice, and the incorporation of English activities into daily life. They had classrooms, outdoor spaces and a schedule as well with vocabulary practice, conversations, afternoon activities and evening review. And the program is helped by the atmosphere,' the coordinator says. You have your students in a collaborative situation so you can wrangle time. They can practice the language (PC) and repeat it with their friends.

Three installation mismatches limited the program ability to flexibly deploy these resources. First, tutor preparation varied. On the comfortable side, two of the tutors employed communicative teaching while as others relied more on explanation, translation, repetition and whole-class control. What had primarily happened in pre-program coordination was scheduling and student management, not shared pedagogy, task sequencing, formative assessment or feedback T3s was quoted saying, 'We have the programme coordination before we do not have any special training' We talk about scheduling and other materials, but not really our teaching or assessment strategies in depth.

Second, items were not grouped in a cohesive module. There were vocabulary lists, worksheets, sample dialogues and topics for conversation but each tutor reused materials from previous cycles while others created their own. Missing task instructions: Specifically, target expressions were not always highlighted; tasks did not define interaction roles (who learns what and through how much other students), expected duration or scaffolding steps (e.g., those used in Task 2) for novel situations and unfamiliar procedures, assessment criteria are unclear on how serendipitous comprehension is to be measured. Third, evaluation and monitoring tools were deficient. There were attendance records and ad hoc (non-standardized) notes from the tutors, but no standardized diagnostic speaking task, weekly progress checklist, or student self-assessment form nor final speaking rubric.

The Installation stage was found to be somewhat inconsistent. Although the environment, facilities, schedule and institutional support were functional, tutor training was less so than structured modules, common rubrics and monitoring tools. These shortcomings impacted process conformity and proof of product later on.

Process Stage: Engaging Participation but Inconsistent Implementation Fidelity

Observed sessions included vocabulary competitions, pair dialogues, pronunciation practice, role plays, short speeches, and group games. Activities involving movement, peer support, humor, and competition generated visible engagement. Students were more willing to speak when mistakes were treated as part of learning and tutors used encouraging language. One student stated, 'Before joining the course, I was afraid to speak English. Here, my friends also make mistakes, so I feel braver. The tutor says mistakes are okay' (S13). The familiar residential peer group appeared to reduce social distance and supported participation.

Despite these strengths, implementation fidelity varied across sessions and tutors. The first discrepancy concerned English exposure. Indonesian and local languages were frequently used to explain instructions, manage behavior, and clarify meaning. This use was understandable for beginners, but the program had no staged policy indicating when English was expected and when Indonesian could be used strategically. T1 explained, 'If we use full English, some students do not understand. We mix Indonesian and English, but sometimes we use too much Indonesian because we want to save time.' The discrepancy was therefore not the presence of the first language, but the absence of agreed pedagogical boundaries.

The second discrepancy concerned student-centeredness. Some sessions provided pair work and role play, whereas others were dominated by tutor explanation and choral repetition. In the latter sessions, students were behaviorally present but had fewer opportunities to initiate, negotiate, or respond in English. The third discrepancy concerned feedback. Tutors frequently praised students with expressions such as 'good,' 'excellent,' and 'try again.' This supported confidence, but feedback rarely specified what was effective, what needed improvement, or how the student should revise and reuse the language. Corrections were sometimes supplied quickly without guided noticing or a follow-up opportunity to produce the form in another context.

The fourth discrepancy concerned monitoring. Daily activities were implemented, but tutor journals and individual progress records were incomplete. Active students were easily noticed, while quieter students could receive less attention. The program lacked a consistent record of who had improved, who needed additional scaffolding, and which activities were most effective. The process stage was rated as moderately discrepant: classroom activity and affective support were strong, but target-language policy, instructional consistency, feedback quality, and monitoring required standardization.

Product Stage: Positive Affective and Behavioral Outcomes but Weak Measurement

The strongest perceived outcome was speaking confidence. Many students reported that repeated practice reduced their fear of speaking. S5 stated, 'Now I am not too shy. My English is not perfect, but I can introduce myself and answer simple questions.' Increased participation was also visible. At the beginning of the course, a limited number of students volunteered; by the final week, more students participated in short dialogues and responded to tutor questions. Tutor notes and observations also indicated growth in basic vocabulary for greetings, self-introduction, daily routines, likes and dislikes, classroom language, and simple opinions.

The residential schedule also supported learning discipline. Students regularly attended sessions, memorized vocabulary, practiced with peers, and joined review activities. However, the evidence was stronger for confidence, participation, self-introduction, and routine vocabulary than for spontaneous response and pronunciation clarity. This pattern is plausible in a short intensive course: willingness and formulaic performance can improve rapidly, whereas spontaneous interaction, pronunciation, and broader linguistic control usually require longer, systematically sequenced practice.

Table 4. Outcome Summary Based on Triangulated Field Evidence

Outcome indicator	Students showing clear evidence	Students showing limited evidence	Main evidence sources
Speaking confidence	19	5	Student interviews and classroom observations
Basic vocabulary use	20	4	Tutor notes and speaking tasks
Participation in pair work	21	3	Classroom observations and student interviews
Short self-introduction	22	2	Final classroom performance and tutor notes
Spontaneous responses	14	10	Classroom observations and tutor notes
Pronunciation clarity	13	11	Tutor feedback and performance observation
Learning discipline	21	3	Attendance records and observations

The figures in Table 4 are not standardized pre-test-post-test scores. They are evaluative counts derived from the evidence rule described in the method. Clear evidence required convergence across at least two sources; limited evidence indicated one source or inconsistent support. The program could therefore demonstrate meaningful affective and behavioral progress, but it could not make strong claims about language-proficiency gains. The absence of common diagnostic and final tasks and a speaking rubric constituted a major product-stage discrepancy.

Cost-Benefit Stage: Contextually Valuable but Under-Documented

Stakeholders regarded the program as worthwhile because it used existing residential resources to increase English exposure, build confidence, establish routines, and strengthen the institution's language-learning identity. Dormitory routines, peer groups, tutors, and existing learning spaces made the program feasible without requiring an entirely separate institutional structure. The principal discrepancy concerned evidence-based reporting. The coordinator stated, 'Parents often ask whether their children improve. We can say yes because we see the difference, but we need a better report, perhaps with scores or portfolios' (PC). Benefits were recognized, but the program did not consistently document student progress, resource use, tutor reflections, or outcomes in a form that could support accountability and cumulative improvement. The cost-benefit stage was therefore rated as moderately discrepant.

Table 5. Evaluation Matrix and Improvement Priorities

DEM stage	Expected standard	Actual condition	Discrepancy level	Priority action
Design	Measurable outcomes, curriculum map, task sequence, and assessment plan	Strong contextual vision, but incomplete weekly outcomes, sequencing, and criteria	Moderate	Develop an outcome-based curriculum and weekly performance targets
Installation	Prepared tutors, structured modules, rubrics, and monitoring tools	Facilities and schedule were ready; training, modules, rubrics, and progress tools were limited	Moderate	Standardize modules and conduct tutor preparation before each cycle
Process	Communicative, scaffolded, feedback-rich, consistently monitored implementation	Engaging activities occurred, but English use, student-centeredness, feedback, and monitoring varied	Moderate	Adopt a staged English-use policy, observation checklist, and feedback protocol
Product	Measurable and documented affective, behavioral, and linguistic outcomes	Confidence and participation improved, but diagnostic/final speaking evidence was incomplete	Major	Use common speaking tasks, analytic rubrics, self-assessment, and portfolios
Cost-benefit	Documented benefits that justify resources and inform stakeholders	Benefits were perceived but not systematically recorded or reported	Moderate	Produce a concise DEM-based report after each program cycle

Across the five stages, the program's most important pattern was a contrast between strong contextual and motivational foundations and weaker technical systems. The design was meaningful but insufficiently measurable; installation provided time and environment but incomplete pedagogical support; the process was engaging but uneven; products were positive but under-measured; and benefits were plausible but under-reported. Assessment and documentation were the most urgent priorities because the major product discrepancy restricted both program learning and stakeholder accountability. Curriculum mapping and tutor preparation formed the next priorities because they influenced all subsequent stages.

The findings indicate that the Madani Course should not be characterized as an ineffective program. Its principal strength is contextual embeddedness. The course uses residential time, peer familiarity, disciplined routines, moral expectations, and tutor mentorship as learning resources. This interpretation supports recent studies showing that pesantren can create distinctive conditions for English exposure, motivation, language policy, and context-responsive pedagogy [1], [2], [9], [10], [11], [12], [13], [15]. The program is not an imported course placed inside a pesantren; it is a locally organized language-learning arrangement shaped by pesantren culture. However, contextual relevance and activity are not equivalent to alignment. DEM reveals that a program may be active, popular, and motivational while still containing substantial gaps between standards and performance [16], [17], [20]. This finding extends previous pesantren research by showing that environment, institutional support, and motivation are necessary but insufficient. They must be connected to explicit outcomes, resource readiness, implementation fidelity, assessment evidence, and reporting. Similar CIPP evaluations have identified strengths in context and process while recommending clearer competency standards, periodic assessment, and stronger documentation [42], [43]. The present study adds a discrepancy logic that identifies the severity and downstream consequences of each gap.

The design-stage findings demonstrate why a broad objective such as 'improving speaking' is inadequate for curriculum control. Communicative competence requires learners to perform identifiable language functions in socially appropriate ways [21], [22]. For beginner learners, a usable curriculum should

specify what students will be able to do at the end of each week and what evidence will demonstrate that ability. Examples include greeting and introducing oneself, requesting clarification, describing routines, expressing preferences, asking follow-up questions, and completing a guided role play. Such targets allow vocabulary, pronunciation, grammar, interaction, and confidence to be taught as coordinated resources rather than disconnected topics.

Task progression is equally important. The student's movement from a vocabulary lesson directly to a speech illustrates how intensity can be confused with speed. Engagement research emphasizes that meaningful participation depends on tasks that are appropriately challenging, socially supported, and clearly structured [25]. Oral development also requires sustained input and repeated opportunities to notice, rehearse, interact, and perform [26]. A more coherent sequence for Madani Course would move from tutor modeling and comprehension checks to choral or individual rehearsal, pair practice, guided conversation, role play, and increasingly independent performance.

The installation discrepancy concerning tutor readiness explains part of the variation observed in the process stage. Recent studies in pesantren report that teacher background, pedagogical strategy, contextual understanding, and institutional support strongly shape English learning [3]-[6], [10], [13]-[15]. In a residential program, tutors are simultaneously instructors, mentors, motivators, and language models. Preparation should therefore cover shared outcomes, communicative task design, scaffolding, the staged use of English, corrective feedback, assessment rubrics, and progress documentation. Standardization should not eliminate tutor creativity; it should provide a common minimum structure within which creativity remains aligned with program goals.

The supportive climate helps explain the visible growth in confidence and participation. Research consistently links positive classroom environments, peer familiarity, and teacher support with willingness to communicate, while anxiety is negatively associated with language achievement [29], [30]-[35]. The Madani Course reduced the social risk of speaking because students shared routines, observed peers making mistakes, and received encouragement from tutors. Confidence functioned as a developmental bridge from silence to participation. However, confidence alone does not make one an expert. As students may become more willing to speak while relying on committed phrases or simply producing language that stays incomprehensible. It is therefore vital to distinguish between affective readiness vs linguistic performance. Confident participation would not be a separate but legitimate short-term goal, and each of fluency, vocabulary range, pronunciation intelligibility, grammatical control, interaction between interlocutors and task completion must be measured separately. This distinction would act to avoid respective underestimating of affective progress and overclaiming of language gains.

We should also be careful in the way we read news about the continued use of Indonesian and local languages. One way that judicious deployment of the L1 can help beginners is by supporting oral comprehension while decreasing cognitive processing load and mitigating anxiety. The issue was not multilingual help, but rather that of inconsistency. A model could necessitate English when greeting, during classroom routines, for pair-task expressions and simple interactions while permitting Indonesian for complex explanations or when emerging questions/urgent clarification arises. This would make exposure to the target language predictable and age-appropriate instead of requiring an unrealistic policy for the exclusivity of English.

Feedback represented another important process gap. General praise helped establish psychological safety, but it did not consistently help learners close the gap between current and expected performance. Formative-assessment research distinguishes encouragement from information that guides improvement [36]-[40]. Effective feedback in Madani Course should identify the successful feature, specify one manageable improvement, provide or elicit a model, and create an immediate opportunity for the learner to try again. The findings are consistent with evidence that the impact of feedback depends on its informational content and on learners' opportunities to understand and use it [28], [38]-[40].

The major product discrepancy is the most consequential result. Interviews, observations, and tutor notes indicated meaningful improvement, but the lack of common diagnostic and final tasks prevented robust claims about change. A valid speaking assessment should represent the outcomes the course claims to develop [37]). For this program, a practical assessment system could include a short diagnostic conversation in the first week, weekly performance tasks, tutor checklists, student self-reflection, a final comparable speaking task, and an analytic rubric covering task completion, fluency, vocabulary, pronunciation, grammatical control, interaction, and confidence. Portfolios would connect assessment with accountability. A simple portfolio could contain initial and final audio or video samples, weekly vocabulary or reflection records, tutor feedback, attendance, and selected task products. This would allow students and parents to see change, help tutors identify persistent needs, and give management evidence for program decisions. It would also strengthen cost-benefit reporting because benefits could be considered alongside tutor time, facilities, supervision, and opportunity costs. Recent pesantren program evaluations similarly recommend competency standards, periodic assessment, mentor development, and more objective outcome evidence [42]-[44].

The recommended improvement sequence is therefore cumulative. First, define measurable communicative outcomes. Second, align weekly materials and tasks with those outcomes. Third, prepare tutors

to apply a shared communicative and feedback protocol. Fourth, implement a staged target-language policy. Fifth, use diagnostic, formative, and final assessment with portfolios. Sixth, produce a concise DEM-based report after each cycle. This sequence retains the program's cultural and motivational strengths while strengthening its technical infrastructure.

The study's theoretical novelty is the integration of DEM with communicative language learning, formative assessment, and pesantren educational culture. DEM alone identifies standards-performance gaps, but it does not specify the content of a high-quality English course. Communicative pedagogy and assessment supply those content standards, while the pesantren lens prevents the standards from being applied without sensitivity to residential routines, moral education, peer life, and institutional identity. The resulting program-alignment framework connects vision, outcomes, resources, implementation, evidence, and reporting in one evaluative chain. Methodologically, the study contributes an instrument grid, explicit discrepancy rules, and a triangulated matrix that other non-formal or residential English programs can adapt. The framework makes qualitative judgement more transparent by showing what standard was used, what evidence described actual performance, why a discrepancy level was assigned, and what action followed. The descriptive student counts further clarify the distribution of qualitative outcome evidence without misrepresenting the design as a quantitative effectiveness study.

Practically, program managers should prioritize assessment and documentation, followed by curriculum mapping and tutor development. Tutors need concise modules, shared weekly targets, model tasks, feedback prompts, and rubrics. Students need visible progression and opportunities to compare early and later performances. Parents and institutional leaders need brief reports that distinguish participation, confidence, and discipline from demonstrated language performance. These changes would improve both learning and accountability without weakening the pesantren identity that gives the program much of its value. The findings are analytically transferable to intensive English courses that share similar characteristics, particularly short duration, residential routines, peer practice, locally developed materials, and limited assessment systems. They should not be statistically generalized to all pesantren. The study was limited to one program cycle and one institutional context; it did not use standardized pre-test-post-test measures; some product judgements relied on triangulated qualitative evidence; and discrepancy classifications remained interpretive despite explicit criteria, member checking, and peer debriefing. Future research should combine DEM with longitudinal speaking assessment, portfolios, multiple evaluators, inter-rater calibration, student outcome tracking, and comparative analysis across several pesantren programs.

4. CONCLUSION

The Madani Course Program is a contextually relevant intensive English course with strong cultural, motivational, and organizational foundations. Its residential environment, peer interaction, repeated practice, disciplined schedule, and supportive climate increased students' willingness to speak, participation, basic vocabulary use, and learning discipline. The DEM analysis nevertheless showed that these strengths were not fully translated into an aligned technical system. Moderate discrepancies affected design, installation, process, and cost-benefit reporting, while a major discrepancy affected the product stage. General goals had not been consistently operationalized as weekly measurable outcomes; tutor preparation, structured modules, rubrics, and monitoring tools were incomplete; implementation varied in English use, student-centeredness, feedback, and documentation; and perceived progress was not supported by standardized diagnostic and final speaking evidence. The study therefore concludes that the program should be continued with systematic improvement rather than replaced. The principal implication is the adoption of an integrated program-alignment framework: define measurable communicative outcomes, sequence weekly tasks, prepare tutors, establish a staged English-use policy, implement formative and final speaking assessment, maintain student portfolios, and issue a brief DEM-based report after each cycle. This framework translates the program's educational vision into evidence-based and sustainable practice while preserving its pesantren identity.

ACKNOWLEDGEMENTS

This research received no external funding. The authors thank the management, tutors, and students of the Madani Course Program at Pondok Pesantren Nurul Haramain Lombok for their support during data collection.

AUTHOR CONTRIBUTIONS

Conceptualization, M.J. and N.N.P.; methodology, M.J.; validation, N.N.P., I.G.A.L.P.U., and P.S.; formal analysis, M.J.; investigation, M.J.; resources, M.J.; data curation, M.J.; writing-original draft preparation, M.J.; writing-review and editing, N.N.P., I.G.A.L.P.U., and P.S.; supervision, N.N.P., I.G.A.L.P.U., and P.S.; project administration, M.J. All contributors have reviewed and consented to the final version of the document.

CONFLICTS OF INTEREST

The author(s) declare no conflict of interest.

INFORMED CONSENT STATEMENT

Institutional permission was obtained before data collection. Participants were informed about the study purpose, voluntary participation, confidentiality, and academic use of the data. Participant identities were protected through coding. All images in this article have been authorized and are the property of the authors.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the generation, analysis, or writing of this manuscript. All aspects of the research, including data collection, interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

- [1] A. C. Y. Bestari, J. Paramboor, D. Faiza, and M. H. Mayekti, "Language policy in Indonesian Islamic boarding schools (pondok pesantren): Implementation, challenges and impacts," *Cakrawala J. Kaji. Manaj. Pendidik. Islam dan Stud. Sos.*, vol. 10, no. 1, pp. 62–81, 2026, doi: 10.33507/cakrawala.v10i1.3780.
- [2] F. Y. Damayanti and T. Purwaningtyas, "The use of English in an Islamic boarding school in Indonesia: Students' perspective," *J. Islam. Educ. El Madani*, vol. 5, no. 2, pp. 77–84, 2026, doi: 10.55438/jiee.v5i2.214.
- [3] T. N. Fitria, "Integrating English language teaching into Islamic boarding schools: A review of strategy and challenges," *J. English Lang. Pedagog.*, vol. 1, no. 2, pp. 64–78, 2023, doi: 10.51826/jelpa.v1i2.772.
- [4] Muhaemin and F. R. Muin, "Developing English learning materials for Islamic boarding schools in Indonesia," *Indones. TESOL J.*, vol. 6, no. 2, 2024, doi: 10.24256/itj.v6i2.8610.
- [5] N. A. Putri, M. Anugerahwati, and Y. Febrianti, "English language teaching in Islamic boarding school: Navigating the digital revolution," *ELTALL English Lang. Teaching, Appl. Linguist. Lit.*, vol. 6, no. 1, pp. 24–30, 2025, doi: 10.21154/eltall.v6i1.10827.
- [6] D. M. Sari, "English language teaching in pesantren: Challenges and strategies," *Turots J. Pendidik. Islam*, vol. 5, no. 2, pp. 110–123, 2023, doi: 10.51468/jpi.v5i2.174.
- [7] U. Umar, "English language teaching in pesantren in Indonesia: Development and challenges," *JELL (Journal English Lang. Lit. STIBA-IEC Jakarta)*, vol. 7, no. 1, pp. 55–64, 2022, doi: 10.37110/jell.v7i1.143.
- [8] W. A. Zahra, "Quality education through pesantren, madrasahs, and Islamic schools in globalization dynamics," *Indones. J. Islam. Stud.*, vol. 13, no. 3, pp. 1–14, 2025, doi: 10.21070/ijis.v13i3.1776.
- [9] D. Eliawati, Ngadiso, and K. A. Putra, "The position of English in the linguistic schoolscapes in an Indonesian Islamic boarding school," *Indones. J. EFL Linguist.*, vol. 9, no. 1, pp. 247–276, 2024, doi: 10.21462/ijefl.v9i1.787.
- [10] A. Madkur et al., "Context-responsive pedagogy in English language teaching in Indonesian Islamic boarding schools," *J. Educ. Learn.*, vol. 18, no. 3, pp. 626–638, 2024, doi: 10.11591/edulearn.v18i3.21085.
- [11] F. A. Putri and M. S. A. Pranata, "Optimizing the pesantren environment to enhance speaking skills: A case study," *J. English Lang. Educ.*, vol. 10, no. 4, 2025, doi: 10.31004/jele.v10i4.1208.
- [12] R. Riyandani, H. Soepriyanti, K. Aprianto, and U. Waluyo, "Motivational factors contributing to students' participation in the English learning programs: A study at MTs Putri Nurul Haramain Islamic Boarding School Narmada," *J. Ilm. Profesi Pendidik.*, vol. 9, no. 4, pp. 2387–2394, 2024, doi: 10.29303/jipp.v9i4.2645.
- [13] O. W. Saputra and S. Fatimah, "Exploring teacher difficulties and strategies in teaching English at Islamic Boarding School of Nurmanisy Hikmatullah," *J. English Lang. Teach.*, vol. 13, no. 3, 2024, doi: 10.24036/jelt.v13i3.130834.
- [14] U. P. A. Saputra, W. F. N. R. Rusafi, and W. I. M. Rohmana, "Implementing various strategies in teaching English through literature in Islamic boarding school," *J. English Lang. Teaching, Lit. Cult.*, vol. 3, no. 1, pp. 42–48, 2024, doi: 10.53682/jelte.v3i1.8294.
- [15] N. Sholihah, A. W. Ritonga, and A. Desrani, "Environmental and institutional support: A study of English language education in pesantren settings," *JELITA J. English Lang. Teach. Lit.*, vol. 6, no. 1, pp. 33–49, 2024, doi: 10.56185/jelita.v6i1.878.
- [16] J. L. Buttram and R. W. Covert, *The Discrepancy Evaluation Model: A Systematic Approach for the Evaluation of Career Planning and Placement Programs*. Columbus, OH, USA: ERIC, 1977.
- [17] M. M. Provus, *Discrepancy evaluation: For educational program improvement and assessment*. McCutchan Publishing Corporation., 1971.
- [18] N. Fadhilah and Asmanah, "Model discrepancy sebagai evaluasi program pendidikan," *Turots J. Pendidik. Islam*, vol. 5, no. 2, pp. 1108–1117, 2023, doi: 10.51468/jpi.v5i2.404.
- [19] P. S. Mustafa, "Model discrepancy sebagai evaluasi program pendidikan," *Palapa*, vol. 9, no. 1, pp. 182–198, 2021, doi: 10.36088/palapa.v9i1.1067.
- [20] C. L. S. Stufflebeam, D. L. & Coryn, *Evaluation theory, models, and applications*. Jossey-Bass, 2014.
- [21] M. Canale and M. Swain, "Theoretical bases of communicative approaches to second language teaching and testing," *Appl. Linguist.*, vol. 1, no. 1, pp. 1–47, 1980, doi: 10.1093/applin/I.1.1.
- [22] D. H. Hymes, "On communicative competence," in *Sociolinguistics: Selected Readings*, J. B. Pride and J. Holmes, Eds., Harmondsworth, UK: Penguin, 1972, pp. 269–293.
- [23] C. C. M. Goh and A. Burns, *Teaching Speaking: A Holistic Approach*. Cambridge, UK: Cambridge University Press, 2012.
- [24] D. Nunan, *Task-Based Language Teaching*. Cambridge, UK: Cambridge University Press, 2004.
- [25] J. Philp and S. Duchesne, "Exploring engagement in tasks in the language classroom," *Annu. Rev. Appl. Linguist.*, vol.

- 36, pp. 50–72, 2016, doi: 10.1017/S0267190515000094.
- [26] K. Saito and K. Hanzawa, “The role of input in second language oral ability development in foreign language classrooms: A longitudinal study,” *Lang. Teach. Res.*, vol. 22, no. 4, pp. 398–417, 2018, doi: 10.1177/1362168816679030.
- [27] M. Sato, “Interaction Mindsets, Interactional Behaviors, and L2 Development: An Affective-Social-Cognitive Model,” *Lang. Learn.*, vol. 67, no. 2, pp. 249–283, 2017, doi: 10.1111/lang.12214.
- [28] M. Sato and S. Loewen, “Metacognitive instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets,” *Lang. Learn.*, vol. 68, no. 2, pp. 507–545, 2018, doi: 10.1111/lang.12283.
- [29] M. Abrar, A. Mukminin, A. Habibi, F. Asyraf, M. Makmur, and L. Marzulina, “‘If our English isn’t a language, what is it?’ Indonesian EFL student teachers’ challenges speaking English,” *Qual. Rep.*, vol. 23, no. 1, pp. 129–145, 2018, doi: 10.46743/2160-3715/2018.3013.
- [30] E. Botes, J.-M. Dewaele, and S. Greiff, “The foreign language classroom anxiety scale and academic achievement: An overview of the prevailing literature and a meta-analysis,” *J. Psychol. Lang. Learn.*, vol. 2, no. 1, pp. 26–56, 2020, doi: 10.52598/jpll/2/1/3.
- [31] C. Boudreau, P. D. MacIntyre, and J.-M. Dewaele, “Enjoyment and anxiety in second language communication: An idiodynamic approach,” *Stud. Second Lang. Learn. Teach.*, vol. 8, no. 1, pp. 149–170, 2018, doi: 10.14746/ssl.2018.8.1.7.
- [32] J.-M. Dewaele and L. Dewaele, “Are foreign language learners’ enjoyment and anxiety specific to the teacher? An investigation into the dynamics of learners’ classroom emotions,” *Stud. Second Lang. Learn. Teach.*, vol. 10, no. 1, pp. 45–65, 2020, doi: 10.14746/ssl.2020.10.1.3.
- [33] G. H. Khajavy, P. D. MacIntyre, and E. Barabadi, “Role of the emotions and classroom environment in willingness to communicate,” *Stud. Second Lang. Acquis.*, vol. 40, no. 3, pp. 605–624, 2018, doi: 10.1017/S0272263117000304.
- [34] P. D. MacIntyre, R. Clément, Z. Dörnyei, and K. A. Noels, “Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation,” *Mod. Lang. J.*, vol. 82, no. 4, pp. 545–562, 1998, doi: 10.1111/j.1540-4781.1998.tb05543.x.
- [35] Y. Teimouri, J. Goetze, and L. Plonsky, “Second language anxiety and achievement: A meta-analysis,” *Stud. Second Lang. Acquis.*, vol. 41, no. 2, pp. 363–387, 2019, doi: 10.1017/S0272263118000311.
- [36] P. Black and D. Wiliam, “Classroom assessment and pedagogy,” *Assess. Educ. Princ. Policy Pract.*, vol. 25, no. 6, pp. 551–575, 2018, doi: 10.1080/0969594X.2018.1441807.
- [37] H. D. Brown and P. Abeywickrama, *Language Assessment: Principles and Classroom Practices*, 3rd ed. New York, NY, USA: Pearson, 2019.
- [38] D. Carless and D. Boud, “The development of student feedback literacy: enabling uptake of feedback,” *Assess. Eval. High. Educ.*, vol. 43, no. 8, pp. 1315–1325, 2018, doi: 10.1080/02602938.2018.1463354.
- [39] M. Henderson *et al.*, “Conditions that enable effective feedback,” *High. Educ. Res. Dev.*, vol. 38, no. 7, pp. 1401–1416, 2019, doi: 10.1080/07294360.2019.1657807.
- [40] B. Wisniewski, K. Zierer, and J. Hattie, “The power of feedback revisited: A meta-analysis of educational feedback research,” *Front. Psychol.*, vol. 10, 2020, doi: 10.3389/fpsyg.2019.03087.
- [41] W. N. Hidayah and A. Shofiya, “Students’ perceptions of microlearning for English learning in an Islamic boarding school,” *J. English Educ. Progr.*, vol. 7, no. 1, pp. 13–25, 2026, doi: 10.26418/jeep.v7i1.102573.
- [42] N. Fitri, S. Kiram, Suwadi, W. Priyangan, and C. L. Jabil, “Evaluating an Arabic-English bilingual program in an Islamic boarding school through the CIPP model: Strengthening students’ language competence,” *Pendas J. Ilm. Pendidik. Dasar*, vol. 11, no. 2, 2026, doi: 10.23969/jp.v11i02.52692.
- [43] Y. Sari, N. N. Padmadewi, P. Suarcaya, I. G. A. L. P. Utami, and D. P. Ramendra, “The evaluation of integrating English language learning program in Islamic boarding school: The CIPP model,” *J. Penelit. dan Eval. Pendidik.*, vol. 28, no. 2, pp. 210–220, 2024, doi: 10.21831/pep.v28i2.75962.
- [44] Z. A. Jamil, Nurhuda, Alfian, S. M. Ulfah, Rasidin, and A. Youssef, “Program evaluation study on Islamic religious education in pesantren: Addressing educational degradation in the digital era,” *J. Pendidik. Agama Islam*, vol. 22, no. 1, pp. 122–139, 2025, doi: 10.14421/jpai.v22i1.11358.
- [45] S. Campbell *et al.*, “Purposive sampling: Complex or simple? Research case examples,” *J. Res. Nurs.*, vol. 25, no. 8, pp. 652–661, 2020, doi: 10.1177/1744987120927206.
- [46] M. Hennink and B. N. Kaiser, “Sample sizes for saturation in qualitative research: A systematic review of empirical tests,” *Soc. Sci. Med.*, vol. 292, 2022, doi: 10.1016/j.socscimed.2021.114523.
- [47] K. Malterud, V. D. Siersma, and A. D. Guassora, “Sample size in qualitative interview studies: Guided by information power,” *Qual. Health Res.*, vol. 26, no. 13, pp. 1753–1760, 2016, doi: 10.1177/1049732315617444.
- [48] K. Vasileiou, J. Barnett, S. Thorpe, and T. Young, “Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period,” *BMC Med. Res. Methodol.*, vol. 18, 2018, doi: 10.1186/s12874-018-0594-7.
- [49] D. Byrne, “A worked example of Braun and Clarke’s approach to reflexive thematic analysis,” *Qual. Quant.*, vol. 56, no. 3, pp. 1391–1412, 2022, doi: 10.1007/s11135-021-01182-y.
- [50] G. Guest, E. Namey, and M. Chen, “A simple method to assess and report thematic saturation in qualitative research,” *PLoS One*, vol. 15, no. 5, 2020, doi: 10.1371/journal.pone.0232076.
- [51] M. E. Kiger and L. Varpio, “Thematic analysis of qualitative data: AMEE Guide No. 131,” *Med. Teach.*, vol. 42, no. 8, pp. 846–854, 2020, doi: 10.1080/0142159X.2020.1755030.
- [52] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, “Thematic analysis: Striving to meet the trustworthiness criteria,” *Int. J. Qual. Methods*, vol. 16, pp. 1–13, 2017, doi: 10.1177/1609406917733847.
- [53] Y. S. Lincoln and E. G. Guba, *Naturalistic Inquiry*. Beverly Hills, CA, USA: SAGE, 1985.