Journal of Education and Teaching
Vol. 1, No.1 (2026): P.30-43
https://journal.terbaik.or.id/index.php/JETINDO

Probing-Prompting Questioning for Fifth-Grade Science Learning: Evidence from a Two-Cycle Classroom Action Study in Indonesia

Candriyani Pagune
Primary School Teacher Education Program, Faculty of Teacher Training and Education, Universitas Pohuwato, Gorontalo, Indonesia
***Corresponding author:** CandriyaniPagune@gmail.com

Article History	Received: 11/04/2026	Accepted: 10/05/2026	Published: 20/05/2026

ABSTRACT

Teacher-centred recitation can restrict pupils' opportunities to explain, revise, and connect scientific ideas. This study examined learning changes during a probing-prompting questioning intervention in a fifth-grade integrated science and social studies class in Gorontalo Province, Indonesia. Collaborative classroom action research involved 20 pupils, including 10 girls and 10 boys, across a pre-intervention stage and two action-reflection cycles. Achievement records and structured activity observations provided the quantitative evidence. An exploratory within-pupil reanalysis used the Friedman test, exact permutation Wilcoxon comparisons with Holm adjustment, Kendall's W, Cochran's Q, exact McNemar tests, and normalized gain. Mean achievement rose from 50.00 (SD = 17.77) before the intervention to 67.55 (SD = 15.93) in Cycle I and 82.50 (SD = 7.86) in Cycle II. The stage difference was substantial, $\chi^2(2) = 30.70$, $p < .001$, $W = .77$, and the pre-intervention to Cycle II normalized gain was .65. Mastery increased from 15% to 55% and 90%, $Q(2) = 22.53$, $p < .001$, while observed activity increased descriptively from 46.25% to 60.95% and 78.40%. Probing-prompting appears useful as a dialogic and formative scaffold for making pupil thinking visible and refining instruction. However, the one-class design, non-equivalent stage measures, short duration, and unreported instrument reliability preclude a strong causal claim. Controlled studies should use validated parallel tests, independent observation, retention measures, and fidelity checks.

Keywords: *classroom action research; dialogic teaching; formative assessment; primary science; probing-prompting; teacher questioning*

1. INTRODUCTION

Primary science learning requires more than the accurate recall of isolated facts. Pupils must connect prior experience with new evidence, express tentative explanations, evaluate alternative ideas, and revise their reasoning. These demands become especially important when pupils study invisible or abstract processes such as gas exchange and respiratory mechanisms. Guided inquiry research in primary science shows that teachers can support critical thinking when they invite pupils to construct, review, and refine representations rather than merely reproduce a final explanation (Kirk et al., 2023; Osborne, 2019). A classroom that limits talk to short, teacher-approved answers therefore risks concealing misconceptions and reducing the intellectual work that pupils perform. Productive science instruction needs an interactional structure that exposes thinking while keeping conceptual goals visible.

Research from Indonesia has reported promising results for probing-prompting across school subjects, although most studies rely on small samples or local action-research designs. Anisah and Carlian (2020) reported improved critical thinking, while Fembriani and Gewahi (2021) reported higher fifth-grade science achievement after a one-group probing-prompting intervention. Manoppo et al. (2022) found achievement differences in a mathematics study, while Jannah et al. (2023) documented increased science concept understanding across two classroom-action cycles. Usman et al. (2023) also reported improvement across two cycles in university foreign-language vocabulary learning. Together, these studies indicate broad feasibility, but they leave three needs unresolved: more transparent reporting of intervention components, stronger analysis of individual repeated measures, and more explicit treatment of design limitations.

The present study addressed a practical problem in one fifth-grade classroom where initial observation indicated low participation, reliance on teacher-centred explanation, and weak science achievement. The original classroom action research aimed to improve learning through iterative implementation and reflection. This article extends that work by reconstructing the intervention for an international audience, anonymising pupil-level reporting, and conducting an exploratory reanalysis of the recorded individual scores. The study asked: (1) How did achievement and mastery change across the pre-intervention stage, Cycle I, and Cycle II? (2) How did observed classroom activity change across the same stages? (3) Which pedagogical mechanisms and implementation refinements plausibly explain the observed trajectory? The article treats the evidence as practice-based and developmental, not as a controlled estimate of causal impact.

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

2. LITERATURE REVIEW

Dialogic teaching provides a useful theoretical frame for such interaction. Dialogic pedagogy treats classroom talk as a medium for joint meaning making, not simply as a delivery channel for teacher knowledge. Reviews identify open invitations, elaboration, uptake of pupil contributions, reasoned challenge, and cumulative development of ideas as central features of productive dialogue (Alexander, 2018; Cui & Teo, 2021; Laird-Gentle et al., 2023; Mercer et al., 2009). Large-scale classroom evidence also links specific forms of teacher-pupil dialogue with learning when teachers encourage pupils to explain, compare, and build on ideas (Hennessy et al., 2023; Howe et al., 2019). These findings do not imply that every question creates dialogue. Closed recall questions can preserve a monologic pattern, whereas linked questions that respond to pupil reasoning can create a genuine space for analysis and revision.

Probing-prompting questioning operationalises this distinction through two complementary moves. A probe asks a pupil to clarify, justify, connect, or extend an initial answer. A prompt supplies a graduated cue when the pupil cannot yet produce a productive response. The teacher can begin with a phenomenon or visual stimulus, allow time for small-group formulation, nominate a respondent, invite peer appraisal, and then adjust the next question to the quality of the emerging answer. In this sense, the approach combines dialogic participation with contingent scaffolding. A recent K-12 review describes effective scaffolding as adaptive support that responds to learner need and gradually transfers responsibility to the learner (Dominguez & Svihla, 2023). Question-chain research similarly shows that the sequence and cognitive function of linked questions matter more than question frequency alone (Wang et al., 2024).

The same questioning sequence can also function as formative assessment. Each answer gives the teacher evidence about current understanding, while a prompt or follow-up probe converts that evidence into an instructional response. This feedback loop aligns with a contemporary meta-analysis of 118 K-12 studies, which found a positive overall association between formative assessment and academic achievement across subjects and settings (Yao et al., 2024). The expected mechanism therefore does not rest on questioning as an isolated technique. It rests on a cycle in which the teacher elicits thinking, interprets the response, provides contingent support, invites further reasoning, and uses the revised response to decide what should happen next.

3. RESEARCH METHOD

3.1 Research Design

The study used collaborative classroom action research organised around planning, action, observation, and reflection. The design followed the cyclical logic associated with Kemmis and McTaggart, in which evidence from one cycle informs changes in the next cycle. The design suited the study's primary purpose because the researcher sought to improve an identified classroom practice while documenting changes in pupil learning. The design did not include random assignment, a comparison class, or a counterfactual condition. Consequently, the study can demonstrate a temporally ordered pattern of change within one classroom, but it cannot isolate the intervention from maturation, repeated testing, teacher attention, or other concurrent influences (Kemmis et al., 2014).

3.2 Setting and Participants

The study took place from November to December 2025 in a public primary school in Tilamuta, Boalemo Regency, Gorontalo Province, Indonesia. The focal class comprised all 20 fifth-grade pupils, with an equal distribution of 10 girls and 10 boys. The school set a minimum mastery criterion of 75 on a 0 to 100 scale and defined class-level success as at least 85% of pupils reaching that criterion. The researcher selected the intact class because initial interviews and observations identified low participation and low attainment in the integrated science and social studies subject, locally termed IPAS. The intervention focused on the science content of human respiration within the unit on living and growing.

3.3 Intervention

The probing-prompting intervention combined a stable sequence of dialogic moves with cycle-specific refinements. The teacher first presented a new situation, image, or respiratory-system problem. Pupils received thinking time and discussed possible answers in small groups. The teacher then posed a whole-class question linked to the lesson objective and nominated a pupil to respond. When an answer was accurate, the teacher asked peers to confirm, challenge, or extend it. When an answer was incomplete or absent, the teacher supplied smaller prompts that pointed toward relevant relationships without immediately giving the final answer. The sequence ended with a higher-demand question and a teacher-supported synthesis. This structure aimed to increase participation while maintaining conceptual direction. Table 1 summarises the stable components and Cycle II refinements.

Table 1
Core Intervention Components and Cycle II Refinements

Phase	Core practice	Refinement after Cycle I
Elicit	Present a visual, phenomenon, or problem and activate prior knowledge.	Restate the learning purpose and reconnect questions explicitly to prior material.
Formulate	Provide thinking time and small-group discussion before public response.	Use more structured question order and additional guidance for pupils who struggled in Cycle I.
Probe	Ask a pupil to explain, justify, connect, or elaborate an initial answer.	Increase follow-up questions and invite more peer responses to incomplete explanations.

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

Phase	Core practice	Refinement after Cycle I
Prompt	Offer graduated cues when a pupil cannot yet produce an adequate explanation.	Use clearer intermediate cues and reinforce correct conceptual links before advancing.
Consolidate	Invite peer evaluation, summarise the reasoning, and complete an individual assessment.	Add guided evaluation, explicit feedback, reflection, motivation, and further practice.

Note. The table reconstructs the procedure reported in the thesis. The source did not report lesson duration, the number of questions per phase, or a formal fidelity score.

3.4 Measures

The primary outcome consisted of stage-level achievement scores recorded for each pupil on a 0 to 100 scale. The pre-intervention assessment contained five multiple-choice questions about respiratory organs, gases, breathing mechanisms, filtration, and respiratory illness. The source document also described five post-test items that sampled related respiratory concepts. The Cycle I score incorporated the post-test and a pupil worksheet, which explains the non-20-point score increments in that stage. The Cycle II table reported a final post-test score. Because the source did not provide item-level responses, psychometric evidence, or a formal parallel-form analysis, this article treats the three scores as stage-level achievement indices rather than strictly equivalent measurements.

Structured observations supplied the secondary evidence. The pre-intervention form recorded 12 general activity indicators, including attention, participation, questioning, group discussion, expression of ideas, task completion, and lesson closure. The cycle form recorded 17 indicators that aligned more closely with the probing-prompting sequence. Observers scored each indicator dichotomously and converted the total to a percentage. The study also recorded affective and psychomotor ratings on four-point categories. Because the pre-intervention and cycle activity forms contained different item sets, the article reports activity percentages descriptively and does not subject them to inferential repeated-measures tests.

3.5 Research Procedures

The researcher administered the pre-intervention assessment on 12 November 2025 after a conventional lesson. Cycle I began on 13 November 2025 and introduced the full probing-prompting sequence. Reflection on Cycle I identified uneven participation, inaccurate responses, and difficulty with the end-of-cycle assessment. The researcher and class teacher then revised the sequence for Cycle II by ordering the questions more explicitly, adding intermediate prompts, increasing feedback, guiding the final evaluation, and reinforcing motivation. Planning for Cycle II occurred on 17 November 2025, after which the class completed the refined lesson and final assessment. The researcher delivered the instruction in collaboration with the class teacher. Parts of the source document also describe the researcher as an observer, so this article cannot assume independent outcome observation.

3.6 Data Analysis

The original study calculated score percentages and mastery rates. The present article retained those results and added an exploratory within-pupil reanalysis from the 20 recorded score trajectories. Descriptive analysis included the mean, standard deviation, median, interquartile range, range, and 95% confidence interval for each stage. The Friedman test assessed the overall repeated-stage difference, and Kendall's W quantified concordance as an omnibus effect size. Exact sign-permutation versions of the Wilcoxon signed-rank test compared pairs of stages, while the Holm procedure controlled family-wise error. Matched-pairs rank-biserial correlation described pairwise effect magnitude. The analysis also calculated the aggregate normalized gain from the pre-intervention mean to the Cycle II mean.

A separate analysis examined the binary mastery indicator at the school cut-off of 75. Cochran's Q tested the overall change across three stages, and exact McNemar tests with Holm adjustment compared stage pairs. Wilson intervals quantified uncertainty around each mastery proportion. These tests strengthen description of the recorded trajectory, but they do not correct the design's lack of a comparison group or the stage measures' limited equivalence. The article therefore uses statistical significance to describe within-class consistency, not to claim population-level causal effectiveness.

4. RESULTS

4.1 Achievement Trajectory

Achievement increased at each stage and became less dispersed by Cycle II. The class mean rose from 50.00 (SD = 17.77) before the intervention to 67.55 (SD = 15.93) in Cycle I and 82.50 (SD = 7.86) in Cycle II. The pre-intervention to Cycle II mean gain equalled 32.50 points, and every pupil recorded a higher Cycle II score than the pre-intervention score. The aggregate normalized gain equalled .65. The reduction in standard deviation from 17.77 to 7.86 suggests that scores clustered more tightly near the target after the second cycle, although the restricted Cycle II score range and different assessment composition may partly explain that convergence. Table 2 and Figure 1 summarise the full trajectory.

Table 2
Achievement, Mastery, and Observed Activity Across Stages

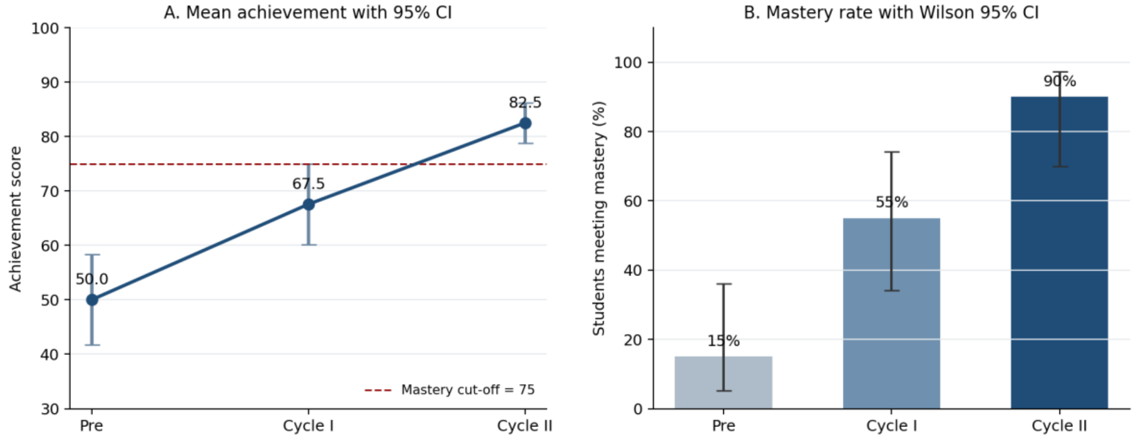
Stage	Achievement M (SD)	Median [IQR]	95% CI for M	Mastery n (%)	Activity M (SD)
Pre-intervention	50.00 (17.77)	40 [20]	41.68 to 58.32	3 (15%)	46.25 (14.94)
Cycle I	67.55 (15.93)	75 [19]	60.10 to 75.00	11 (55%)	60.95 (15.83)
Cycle II	82.50 (7.86)	80 [10]	78.82 to 86.18	18 (90%)	78.40 (12.02)

Note. N = 20 at every stage. Mastery required a score of at least 75. Achievement intervals use the t distribution. Activity values remain descriptive because the pre-intervention and cycle observation forms contained different indicators.

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

Figure 1

Achievement and Mastery Across the Pre-Intervention Stage and Two Action Cycles



Note. Panel A displays t-based 95% confidence intervals around the mean. Panel B displays Wilson 95% confidence intervals around the mastery proportion.

The omnibus analysis confirmed that the recorded achievement ranks differed across stages, Friedman $\chi^2(2) = 30.70$, $p < .001$, Kendall's $W = .77$. Exact pairwise comparisons remained significant after Holm adjustment. The pre-intervention to Cycle I comparison produced a 17.55-point mean gain and a rank-biserial effect of .75. The Cycle I to Cycle II comparison produced a 14.95-point gain and a rank-biserial effect of .96. The full pre-intervention to Cycle II comparison produced a 32.50-point gain, with all 20 paired differences in the positive direction and a rank-biserial effect of 1.00. These results indicate a highly consistent within-class trajectory, although measurement and design limitations still constrain causal interpretation.

Table 3
Exploratory Repeated-Measure Comparisons

Outcome or comparison	Change	Test statistic	Effect size	Holm-adjusted p	Interpretation
Achievement across three stages	Overall	$\chi^2(2) = 30.70$	W = .77	< .001	Large, ordered stage difference
Pre to Cycle I	+17.55 points	W− = 26.5	r_rb = .75	.002	17 improved, 3 declined
Cycle I to Cycle II	+14.95 points	W− = 4.0	r_rb = .96	< .001	18 improved, 2 declined
Pre to Cycle II	+32.50 points	W− = 0.0	r_rb = 1.00	< .001	All 20 pupils improved
Mastery across three stages	+75 percentage points	Q(2) = 22.53	Not applicable	< .001	15%, 55%, and 90%

Note. Pairwise achievement tests used exact sign permutation for tied ranks. r_rb denotes matched-pairs rank-biserial correlation. The mastery row reports Cochran's Q. Exact McNemar comparisons yielded Holm-adjusted p = .016 for pre-intervention versus Cycle I, p = .016 for Cycle I versus Cycle II, and p < .001 for pre-intervention versus Cycle II.

4.2 Mastery Attainment

Mastery increased from 3 of 20 pupils before the intervention to 11 pupils in Cycle I and 18 pupils in Cycle II. The corresponding proportions were 15% (Wilson 95% CI [5.2%, 36.0%]), 55% (95% CI [34.2%, 74.2%]), and 90% (95% CI [69.9%, 97.2%]). Cochran's Q showed an overall stage difference, $Q(2) = 22.53$, $p < .001$. Eight pupils moved from non-mastery to mastery between the pre-intervention stage and Cycle I, and seven additional pupils made that transition between Cycle I and Cycle II. No pupil moved from mastery to non-mastery in the binary stage comparisons. The final 90% rate exceeded the school's class-success criterion of 85%.

4.3 Observed Activity and Broader Learning Indicators

Observed activity followed the same descriptive direction as achievement. Mean activity increased from 46.25% (SD = 14.94) before the intervention to 60.95% (SD = 15.83) in Cycle I and 78.40% (SD = 12.02) in Cycle II. Cycle II affective ratings placed 18 pupils, or 90%, in the good or very good categories and two pupils in the adequate category. Cycle II psychomotor ratings placed 16 pupils, or 80%, in the good or very good categories and four pupils in the adequate category. These observations support the interpretation that the refined questioning cycle increased participation and performance, but the lack of independent raters, inter-rater reliability, and identical observation items prevents a precise estimate of activity growth.

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

Journal of Education and Teaching
Vol.1, No.1 Mei (2026): P.30-43

5. DISCUSSION

5.1 Principal Findings

The study documented a strong and orderly improvement in one classroom during two cycles of probing-prompting questioning. Achievement increased by 32.50 points from the pre-intervention stage to Cycle II, every pupil improved across that full interval, and mastery increased by 75 percentage points. The second cycle also reduced score dispersion and moved the class beyond the school's 85% mastery target. These patterns answer the first research question affirmatively at the classroom level: recorded achievement and mastery improved during the intervention period. The same conclusion requires careful wording at the causal level because the design did not establish what would have happened without the intervention.

The large Kendall's W and positive rank-biserial effects show that the improvement did not depend on only a few high-performing pupils. Instead, most pupils moved in the same direction, and all 20 pupils scored higher in Cycle II than at baseline. This distributional evidence strengthens the original thesis, which relied mainly on mastery percentages. The narrowing score distribution also suggests that the intervention may have supported pupils who began below the mastery threshold. Nevertheless, the Cycle II score range compressed between 60 and 90, so ceiling constraints and assessment construction may also have reduced variance. The pattern supports an equity-oriented interpretation, but the measurement design cannot prove it.

5.2 Why Probing and Prompting May Support Science Learning

The first plausible mechanism involves diagnostic visibility. In a teacher-centred lesson, a correct answer from one pupil can conceal widespread misunderstanding. The intervention repeatedly required pupils to formulate, state, justify, and revise answers. Those responses gave the teacher immediate evidence about which conceptual link had failed, such as confusing lungs with alveoli or misidentifying the muscles involved in breathing. The teacher could then supply a prompt that targeted the missing link. This sequence reflects formative assessment because it connects elicitation directly to instructional adjustment. Yao et al. (2024) found a positive overall effect of formative assessment on K-12 achievement, and the current trajectory illustrates how questioning can embed that process within ordinary instruction.

The second mechanism involves contingent scaffolding. A prompt reduces task difficulty without removing the pupil's responsibility to reason, while a probe increases cognitive demand once an initial response becomes available. This alternation can keep pupils within a productive challenge range. Dominguez and Svihla (2023) argue that teacher-implemented scaffolding becomes most useful when the support responds to learner need rather than following a fixed script. The Cycle II refinements moved the intervention in that direction by adding intermediate cues, ordering questions more deliberately, and responding more explicitly to incorrect answers. The larger Cycle I to Cycle II rank-biserial effect therefore fits a scaffolding account, although the study did not code prompt contingency directly.

The third mechanism involves cumulative dialogue. The intervention asked peers to evaluate a response and allowed groups to formulate answers before whole-class nomination. These moves can shift classroom talk from isolated teacher-pupil exchanges toward collective reasoning. Reviews of dialogic education emphasise that productive dialogue gives pupils opportunities to elaborate, compare perspectives, and build on previous turns (Cui & Teo, 2021; Laird-Gentle et al., 2023). Hennessy et al. (2023) similarly show that productive teacher-pupil dialogue depends on how contributions connect across turns. The present intervention did not transcribe talk, but its structure created more opportunities for those connections than the initial recitation pattern.

The fourth mechanism concerns the architecture of question chains. The teacher did not simply ask more questions. The procedure linked an initial problem, a pupil explanation, a peer response, a targeted prompt, a deeper probe, and a concluding synthesis. Wang et al. (2024) show that question-chain patterns can shape learning because each question performs a different epistemic function. In primary science, this sequence can help pupils move from naming an organ to explaining a process and then applying the relationship to health behaviour. Kirk et al. (2023) also demonstrate that guided primary science inquiry supports critical thinking when pupils construct and refine representations. The use of respiratory-system images and explanation prompts in this study aligns with that representational pathway.

5.3 Why Cycle II Outperformed Cycle I

Action research depends on purposeful refinement, and the Cycle II changes appear instructionally meaningful. Cycle I revealed three barriers: some pupils remained reluctant to respond, several answers lacked accuracy, and many pupils struggled with the final assessment. Cycle II addressed those barriers through more structured question order, clearer prompts, additional feedback, guided evaluation, and motivational reinforcement. The achievement mean increased by another 14.95 points, while seven additional pupils reached mastery. Activity also rose descriptively by 17.45 percentage points. The co-movement of participation and attainment supports the interpretation that improved implementation, not simple exposure to the label “probing-prompting,” drove the second-cycle gains.

This interpretation also explains why the intervention should not become a rigid nine-step routine. Effective dialogue requires the teacher to listen to the substance of a pupil’s answer and decide whether to probe, prompt, redirect, invite peer reasoning, or consolidate. Cui and Teo (2023) argue that dialogue can develop critical thinking when classroom talk makes reasoning itself an object of attention. A fixed sequence that rewards only the expected answer could reproduce recitation under a new name. The practical contribution of the present study therefore lies in its adaptive cycle: diagnose, cue, probe, compare, and consolidate.

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

Journal of Education and Teaching
Vol.1, No.1 Mei (2026): P.30-43

5.4 Relationship to Previous Probing-Prompting Research

The results converge with several Indonesian studies while adding a more cautious analytical interpretation. Fembriani and Gewahi (2021) reported higher fifth-grade science outcomes after probing-prompting, and Jannah et al. (2023) found that concept understanding increased from Cycle I to Cycle II among 19 fourth-grade pupils. Manoppo et al. (2022) identified an achievement advantage in mathematics, while Maryam et al. (2024) reported classroom-action improvement in mathematics. Usman et al. (2023) documented a similar two-cycle increase in foreign-language vocabulary. The recurrence of improvement across domains suggests that the approach offers a general interactional scaffold rather than a subject-specific trick.

At the same time, the shared methodological profile of these studies limits the strength of cumulative evidence. Many studies use one group, small samples, researcher-developed assessments, and short cycles. Such designs serve local improvement well, but they cannot separate intervention effects from repeated exposure, teacher enthusiasm, regression to the mean, or assessment alignment. The present article improves transparency by reporting uncertainty, matched-pupil effects, assessment non-equivalence, and observer limitations. It does not eliminate those limitations. The literature now needs multisite and controlled studies that preserve the adaptive dialogic character of probing-prompting while strengthening causal identification.

5.5 Practical Implications

The findings support a practical design principle for primary teachers: plan the logic of a question chain, not only a list of questions. A useful chain begins with a phenomenon that pupils can inspect, asks for an initial explanation, identifies the specific gap in that explanation, supplies the smallest helpful cue, and then asks the pupil or a peer to reconstruct the explanation. Teachers should allow sufficient thinking time, distribute participation across the class, and treat incorrect answers as diagnostic resources. This design maintains intellectual challenge while reducing the public risk that can make pupils avoid participation.

Teachers should also integrate assessment with dialogue. Short exit questions or parallel items can show whether pupils transfer the reasoning beyond the discussed example. A simple lesson record can document the target concept, the planned probe, likely misconceptions, graduated prompts, peer-response moves, and evidence that signals readiness to advance. Such documentation would improve instructional consistency and create a fidelity record for future research. Schools that adopt the approach should therefore develop both questioning skill and assessment literacy, rather than promoting the model as a script.

5.6 Limitations and Future Research

The study has six limitations that materially shape interpretation. First, one intact class of 20 pupils limits statistical precision and external validity. Second, the absence of a comparison group prevents a causal estimate. Third, the three achievement indices did not use demonstrably equivalent forms, and Cycle I combined a post-test with a worksheet. Fourth, the source reported no item validity, internal consistency, scorer reliability, or delayed retention test. Fifth, the activity instruments changed between baseline and the cycles, and observer independence remained unclear. Sixth, the source did not specify lesson duration, question frequency, prompt fidelity, institutional ethics review, or written consent and assent procedures. These limitations mean that the statistical results describe the recorded classroom trajectory more strongly than they establish intervention effectiveness.

Future research should use a cluster or class-level comparison, validated parallel achievement forms, and a pre-specified analysis plan. Researchers should record lessons and code question functions, wait time, prompt contingency, pupil elaboration, peer uptake, and participation equity. Independent observers should report inter-rater agreement, and a delayed test should examine retention. A mediation design could then test whether increases in elaborated pupil talk or formative feedback explain achievement growth. Researchers should also examine differential participation by prior attainment and gender, because a whole-class mean can conceal who receives probes, who receives prompts, and whose ideas shape the final explanation.

6. CONCLUSION

This two-cycle classroom action study shows that probing-prompting questioning can organise a more participatory and diagnostically responsive fifth-grade science lesson. The class recorded a substantial achievement increase, a rise in mastery from 15% to 90%, and a parallel descriptive increase in classroom activity. The Cycle II refinements suggest that the quality and contingency of the question sequence matter more than simple repetition. Probing made pupil reasoning visible, prompting supplied graduated support, peer responses widened the dialogue, and teacher synthesis stabilised the target concept.

The evidence warrants a practice-based conclusion rather than a definitive causal claim. The approach appears promising for local instructional improvement, especially when teachers use pupil responses to adapt subsequent questions. International publication and wider adoption require stronger ethical documentation, equivalent and validated assessments, independent observation, fidelity evidence, comparison conditions, and retention data. With those safeguards, future studies can determine whether probing-prompting produces durable science learning and more equitable participation across classrooms.

ACKNOWLEDGEMENTS

The author acknowledges the participating school, class teacher, supervisors, pupils, and families who supported the classroom project. Names should be added only after institutional and consent requirements have been verified.

E-ISSN: xxxx-xxxx
P-ISSN: xxxx-xxxx
DOI :

AUTHOR CONTRIBUTIONS

Candriyani Pagune: conceptualisation, methodology, investigation, data curation, formal analysis, visualisation, and writing of the original draft. The final statement must identify any additional journal co-authors and follow the journal's authorship policy.

FUNDING

The source thesis reports no external funding. The author must confirm this statement before submission.

DATA AVAILABILITY

The article derives its aggregate and exploratory statistics from the pupil-level score tables contained in the author's undergraduate thesis. A de-identified dataset and analysis file should accompany submission when the target journal permits supplementary material.

CONFLICTS OF INTEREST

The author declares no conflicts of interest. The author must confirm this statement before submission.

GENERATIVE ARTIFICIAL INTELLIGENCE DISCLOSURE

OpenAI's ChatGPT was used to support English-language editing, manuscript restructuring, and computational checking of descriptive and non-parametric analyses derived from the thesis data. The author reviewed the generated content and remains responsible for the accuracy of the data, interpretation, citations, and research integrity.

REFERENCES

Alexander, R. J. (2018). Developing dialogic teaching: Genesis, process, trial. Research Papers in Education, 33(5), 561–598. https://doi.org/10.1080/02671522.2018.1481140

Anisah, T., & Carlian, Y. (2020). Penerapan model pembelajaran probing prompting untuk meningkatkan kemampuan berpikir kritis siswa madrasah ibtidaiyah. Al-Aulad: Journal of Islamic Primary Education, 3(2), 98–106. https://doi.org/10.15575/al-aulad.v3i2.9047

Cui, R., & Teo, P. (2021). Dialogic education for classroom teaching: A critical review. Language and Education, 35(3), 187–203. https://doi.org/10.1080/09500782.2020.1837859

Cui, R., & Teo, P. (2023). Thinking through talk: Using dialogue to develop students' critical thinking. Teaching and Teacher Education, 125, 104068. https://doi.org/10.1016/j.tate.2023.104068

Dominguez, S., & Svihla, V. (2023). A review of teacher implemented scaffolding in K-12. Social Sciences & Humanities Open, 8(1), 100613. https://doi.org/10.1016/j.ssaho.2023.100613

Fembriani, F., & Gewahi, M. (2021). Pengaruh model probing-prompting terhadap hasil belajar IPA peserta didik kelas V. Jurnal Ilmiah Kontekstual, 2(2), 61–68. https://doi.org/10.46772/kontekstual.v2i02.398

Hennessy, S., Calcagni, E., Leung, A., & Mercer, N. (2023). An analysis of the forms of teacher-student dialogue that are most productive for learning. Language and Education, 37(2), 186–211. https://doi.org/10.1080/09500782.2021.1956943

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher-student dialogue during classroom teaching: Does it really impact on student outcomes? Journal of the Learning Sciences, 28(4–5), 462–512. https://doi.org/10.1080/10508406.2019.1573730

Jannah, O. D. N., Fajrie, N., & Kurniati, D. (2023). Kemampuan pemahaman konsep IPA menggunakan penerapan model pembelajaran probing-prompting dengan media permainan kelereng. PENDASI: Jurnal Pendidikan Dasar Indonesia, 7(2), 251–262. https://doi.org/10.23887/jurnal_pendas.v7i2.2435

Kemmis, S., McTaggart, R., & Nixon, R. (2014). The action research planner: Doing critical participatory action research. Springer. https://doi.org/10.1007/978-981-4560-67-2

Kirk, M., Tytler, R., & White, P. (2023). Critical thinking in primary science through a guided inquiry pedagogy: A semiotic perspective. Teachers and Teaching, 29(6), 615–637. https://doi.org/10.1080/13540602.2023.2191181

Laird-Gentle, A., Larkin, K., Kanasa, H., & Grootenboer, P. (2023). Systematic quantitative literature review of the dialogic pedagogy literature. The Australian Journal of Language and Literacy, 46, 29–51. https://doi.org/10.1007/s44020-022-00029-9

Manoppo, R., Pomalato, S. W. D., Zakiyah, S., & Puloo, M. M. L. (2022). Pengaruh model pembelajaran probing prompting terhadap hasil belajar siswa pada materi segiempat. Jambura Journal of Mathematics Education, 3(2), 93–99. https://doi.org/10.34312/jmathedu.v3i2.15604

Maryam, S. M., Usman, U., & Ramadhany, M. (2024). Penerapan model pembelajaran probing prompting untuk meningkatkan hasil belajar matematika peserta didik. Saintifik, 10(1), 60–66. https://doi.org/10.31605/saintifik.v10i1.453

Mercer, N., Dawes, L., & Staarman, J. K. (2009). Dialogic teaching in the primary science classroom. Language and Education, 23(4), 353–369. https://doi.org/10.1080/09500780902954273

Osborne, J. F. (2019). Not "hands on" but "minds on": A response to Furtak and Penuel. Science Education, 103(5), 1280–1283. https://doi.org/10.1002/sce.21543

Usman, M., Adys, H. P., Rosmaladewi, R., & Asnur, M. N. A. (2023). The application of probing prompting learning models in mastering foreign language vocabulary. International Journal of Language Education, 7(4), 702–710. https://doi.org/10.26858/ijole.v7i4.58366

Wang, J., Wang, Y., Moore, Y., Sneed, S., Thacker, B., & Hart, S. (2024). Analyzing the patterns of questioning chains and their intervention on student learning in science teacher preparation. International Journal of Science and Mathematics Education, 22, 809–836. https://doi.org/10.1007/s10763-023-10408-4

Yao, Y., Amos, M., Snider, K., & Brown, T. (2024). The impact of formative assessment on K-12 learning: A meta-analysis. Educational Research and Evaluation, 29(7–8), 452–475. https://doi.org/10.1080/13803611.2024.2363831