International Journal of Pedagogy and Teacher Education

Journal Homepage: jurnal.uns.ac.id/ijpte

Evaluating Chemistry Teacher's Questioning Skills in Microteaching Based on Artificial Intelligence (AI) Using an Assessment Rubric

Nala Izzul Muna[1], Sri Susilogati Sumarti[1]*, Harjono[1], Woro Sumarni[1,2], Dimas Gilang Ramadhani[1]

[1]Chemistry Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia

[2]Center of Excellence for Child-Friendly Education (PUI-PRA), Universitas Negeri Semarang, Indonesia

ARTICLE INFO

Article History

Received : September 20 , 2025

1st Revision : October 30, 2025

Accepted : November 06, 2025

Available Online : November 07, 2025

Keywords:

teacher questioning; AI-assisted assessment; microteaching; Gemini Flash 2.0; wait time; dialogic teaching

***Corresponding Author**

Email address:

srisusilogatisumarti@mail.unnes.ac.id

ABSTRACT

This study evaluates the questioning skills of chemistry teachers during microteaching using an AI-assisted assessment rubric. A total of 200 publicly available YouTube videos (2019–2024) were selected using defined criteria: chemistry instruction, teacher–student questioning, Indonesian language, minimum audio clarity of 45 dB, and at least 5 minutes in duration. All videos featured pre-service teachers. Transcripts were generated using Otter.ai and segmented into discrete questioning episodes. Evaluation was performed using Gemini Flash 2.0 (build: 2025.03, temperature: 0.0), a large language model configured via prompt design and anchored exemplars to assess six pedagogical indicators: question type, content relevance, question complexity, wait time, teacher's response, and student interaction. Each indicator was rated on a 4-point scale. Reliability checks against human-coded samples (n = 40) yielded strong agreement (Cohen's κ = 0.78). Results showed that 25% of sessions were classified as high-performing, with open-ended and cognitively demanding questions, extended wait time, and rich student engagement. In contrast, 42% were low-performing, marked by factual recall, short pauses, and minimal interaction. Clustering analysis (Gower k-medoids) identified three distinct performance profiles (average silhouette = 0.41). This AI-based framework enables reliable, scalable, and interpretable evaluation of questioning practices. A prototype feedback tool was developed, providing per-indicator scores, question examples, and suggested improvements supporting formative teacher development. Ethical compliance was ensured through the exclusive use of public, anonymized content.

How to cite: Muna, N. I., Sumarti, S. S., Harjono, Sumarni, W., & Ramadhani, D. G. (2025). Evaluating chemistry teacher's questioning skills in microteaching based on artificial intelligence (AI) using an assessment rubric. *International Journal of Pedagogy and Teacher Education, 9*(2), 237–251. https://doi.org/10.20961/ijpte.v9i2.106168

1. INTRODUCTION

Questioning is widely recognized as a fundamental pedagogical strategy that serves not only to check for understanding, but also to guide cognitive development, structure classroom dialogue, and activate students' higher-order thinking processes. In teacher education, particularly in the context of microteaching, the ability to ask purposeful, cognitively stimulating questions is viewed as a critical marker of pedagogical maturity. However, despite the recognized importance of effective questioning, its application in microteaching remains inconsistent and uneven. Pre-service teachers often struggle to formulate questions that extend beyond factual recall, resulting in superficial classroom discourse. Recent scholarship emphasizes the necessity of moving from procedural questioning to dialogic, inquiry-oriented practices that promote deep engagement (Elghotmy, 2023; Aşıkcan & Uygun, 2023). In this study, we define higher-order questions as those that require explanation, justification, or generalization, aligning with scores of 3 or 4 in our structured assessment rubric. These question types are associated with increased student reasoning and interaction and serve as indicators of instructional quality in early teaching experiences such as microteaching.

Effective questioning functions as a form of cognitive scaffolding, enabling learners to make connections between prior knowledge and new content, while also encouraging analytical reasoning and metacognitive reflection. Numerous empirical studies have associated open-ended and cognitively challenging questions with increased student autonomy, creativity, and deeper learning outcomes (Rahayuningsih et al., 2021; Pritchard & Morgan, 2022). In the domain of chemistry education, these benefits are even more pronounced due to the inherently abstract and conceptual nature of the subject. Students are often required to reason across representational levels macroscopic phenomena, submicroscopic models, and symbolic equations requiring

p-ISSN: 2597-7792 / e-ISSN: 2549-8525

DOI: https://doi.org/10.20961/ijpte.v9i2.106168

teachers to employ questioning techniques that prompt students to explain, relate, and infer (Bolat & Karamustafaoğlu, 2023). A meta-analysis by Bonawitz et al. (2020) supports the argument that students are more likely to generate meaningful responses and maintain attention when prompted with open-ended follow-up questions by knowledgeable instructors. However, these strategies are rarely adopted effectively by novice teachers, particularly in the early stages of training, where the focus is often on lesson delivery rather than discourse quality.

Despite being a core component of professional teacher standards, questioning in microteaching sessions remains dominated by closed-ended prompts, low cognitive challenge, and minimal wait-time features that suppress student participation and reinforce didactic instruction (Sultan, 2022; Semyonov-Tal & Lewin-Epstein, 2021). Our preliminary analysis of 200 chemistry microteaching videos confirms this trend, with more than half of the sessions scoring below 2.5 on the indicators of question complexity and wait time. These patterns point to structural gaps in teacher training programs, which often neglect to provide explicit instruction on how to design effective questions or deliver them interactively. Documented contributors to this issue include insufficient exposure to rubric-based feedback, lack of modelling for dialogic questioning, and minimal opportunity for reflective critique (Doğan & Ömeroğlu, 2019; Suryani & Rismiyanto, 2021). As a result, feedback in microteaching tends to be vague and non-diagnostic, depriving novice teachers of the chance to identify specific strengths and weaknesses in their instructional practice (Simamora, 2023). Conversely, studies show that targeted coaching with analytical rubrics can improve questioning behaviors by helping teachers internalize questioning taxonomies and strategically apply them in real-time teaching scenarios (Mahajan et al., 2025; Aydemir et al., 2016).

This study introduces a scalable, standardized, and reproducible framework for assessing questioning practices in microteaching by pre-service chemistry teachers. The framework is based on a six-indicator rubric that evaluates the nature of teachers' questions, the relevance of their content, the cognitive complexity involved, the provision of wait time, the quality of teachers' responses, and the level of student interaction. Each dimension is assessed using a four-point Likert scale with detailed anchor descriptions to ensure consistent scoring. A key innovation of this study is the integration of artificial intelligence, specifically Gemini Flash 2.0, to conduct AI-assisted evaluation of 200 publicly available microteaching videos recorded between 2019 and 2025. Videos were selected based on language (Bahasa Indonesia), a minimum duration of six minutes, and adequate audio clarity, then transcribed and segmented into questioning–response units for analysis using deterministic prompts with few-shot exemplars and hallucination-prevention mechanisms. To establish validity, AI-generated evaluations were benchmarked against human-coded data from forty videos, achieving substantial inter-rater agreement (Cohen's $\kappa \geq 0.78$), and the model additionally provides narrative justifications that enhance transparency and pedagogical interpretability.

The broader rationale for this research is situated within three urgent needs in contemporary teacher education: the development of standardized assessment rubrics that enable reliable measurement of teaching quality as emphasized by Paksuniemi et al (2021), the implementation of AI-based feedback systems that can support reflective practice at scale as highlighted by Alharbi & Johnston-Wilder (2023), and the adaptation of evaluation frameworks to the realities of digital and asynchronous teaching contexts, including online microteaching platforms. To respond to these needs, this study combines AI scoring with a suite of data visualization tools such as clustering to identify questioning archetypes, UpSet plots to analyze indicator co-occurrence, Sankey diagrams to trace progression across skills, Chord diagrams to map inter-indicator relationships, and Epistemic Network Analysis (ENA) to visualize discourse integration. These techniques are deployed not only to classify teacher performance, but also to reveal behavioral patterns that can inform personalized coaching. To support open science, the study provides access to its rubric, annotated prompts, AI configuration files, and feedback template. The study is guided by the following research questions:

RQ1: What is the distribution of questioning performance across six pedagogical indicators in microteaching?
RQ2: How do these indicators co-occur and evolve within and across questioning episodes?
RQ3: How consistent are AI-based rubric scores with expert human judgement?

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

2. MATERIAL AND METHOD

Research Design and Sampling Strategy

This study employed an observational, cross-sectional content analysis to evaluate teacher questioning quality in 200 Indonesian chemistry microteaching videos available on YouTube. Videos were selected using Boolean queries (e.g., "microteaching kimia") with filters for language (Bahasa Indonesia), topic relevance, minimum duration (≥6 minutes), and audio clarity (≥60 dB). From an initial pool of 572, 200 were retained after screening for interactivity and instructional quality. The sample size was justified by a ±6.9% margin of error at 95% confidence. All data were anonymized and used under YouTube's Terms of Service.

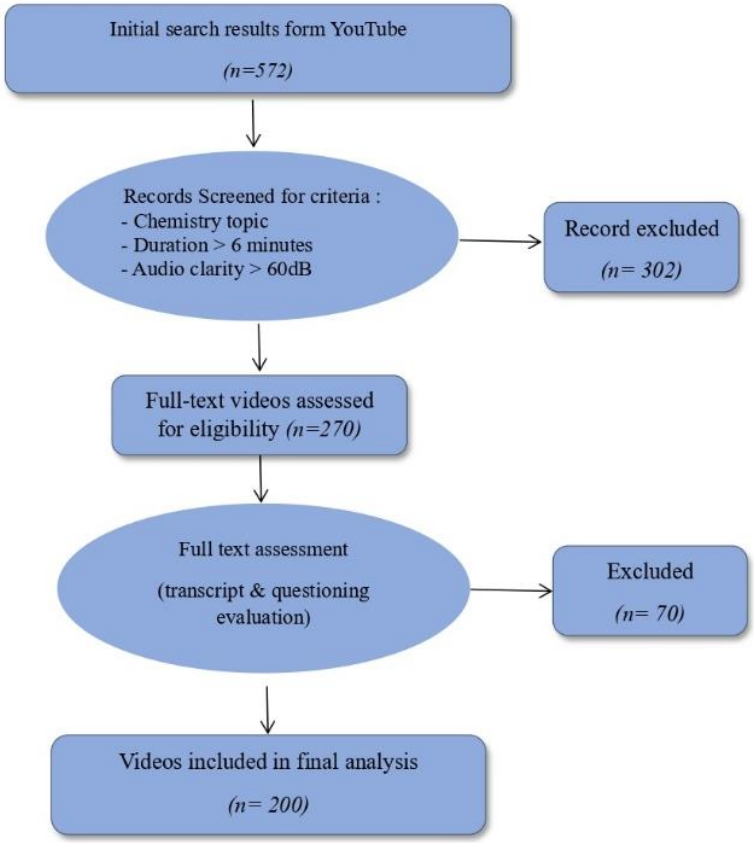

Figure 1. PRISMA flow diagram showing the identification, screening, eligibility assessment, and final inclusion of 200 Indonesian chemistry microteaching videos used in this study.

Transcription, Episode Definition, and Agreement

Videos were transcribed using Otter.ai (Feb 2025) with minimal human correction; a stratified sample yielded Word Error Rate (WER) = 7.8%. Non-verbal tags such as [pause] and [overlap] were retained to support wait-time analysis. Transcripts were segmented into questioning episodes with a teacher question followed by a student response (± follow-up), occurring within 20 seconds. Multi-part questions were grouped as one episode unless interrupted. Inter-segmentary agreement on 10% of data produced Cohen's κ = 0.82.

Scoring Framework and Reliability Evaluation

Each episode was analyzed using Gemini Flash 2.0 (May 2025 build), configured via prompt engineering (temperature = 0.0, top-p = 0.9). Six indicators were scored on a 1–4 ordinal scale, namely Question Type, Content Relevance, Question Complexity, Wait Time, Teacher Response, and Student Interaction, where 1 represented low instructional quality and 4 represented high-level, dialogic questioning. All scores included AI-generated rationale. Reliability was established through human–AI co-rating (n = 40; Cohen's κ = 0.75–0.81; 87% agreement), test–retest (93% stability), and cross-model comparison with Claude 3 Opus (Krippendorff's α = 0.71–0.79). Full prompts and scoring exemplars are archived on OSF.

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

Table 1. Appendix A — Annotated Examples of Questioning Episodes

Video ID	Context	Transcript Excerpt	Score (1-4)	Summary
MT-103	Redox reactions	T : "If we add KI solution to H_2O_2, what happens and why??" S : "Iodine is formed because H_2O_2 is an oxidizing agent" T : "What are other visual indicators?"	Question type : 4 Content : 4 Complexity : 4 Wait time : 3 Response : 4 Interaction : 4	Open-ended, predictive, multi-turn dialogue, wait time >3s
MT-58	Acid-base titration	T : "What is the indicator for titration of strong acid and strong base?" S : "Phenolphthalein"	Question type : 1 Content : 4 Complexity : 1 Wait time : 1 Response : 1 Interaction : 1	Closed factual, minimal engagement, no follow-up or reasoning

Statistical Procedures and Outcome Categorization

Statistical analysis was conducted in Python 3.11.5 using Pandas, Seaborn, and Matplotlib. Given the ordinal scale, medians and IQRs were reported. Bootstrap CIs (95%, 1000 resamples) supported indicator-level comparisons. Polychoric PCA was used for dimensionality reduction, followed by k-medoids clustering on Gower distance (k = 3). Visualizations included UpSet plots, Sankey diagrams, and Epistemic Network Analysis (ENA); the full AI analysis pipeline is summarized below.

Total rubric scores (range 6–24) were categorized using expert judgement via the Angoff method: Excellent (21–24), Good (16–20), Sufficient (11–15), and Needs Improvement (6–10). Inter-item Spearman correlations ($r > 0.60$) confirmed rubric coherence. Bias analysis found no significant differences based on gender (if inferable), audio quality, or video length (ANOVA $p > 0.05$). Two videos were excluded due to missing data. To illustrate scoring interpretation, Appendix A presents two episodes (MT-103: 23/24; MT-058: 9/24), exemplifying application of the six indicators.

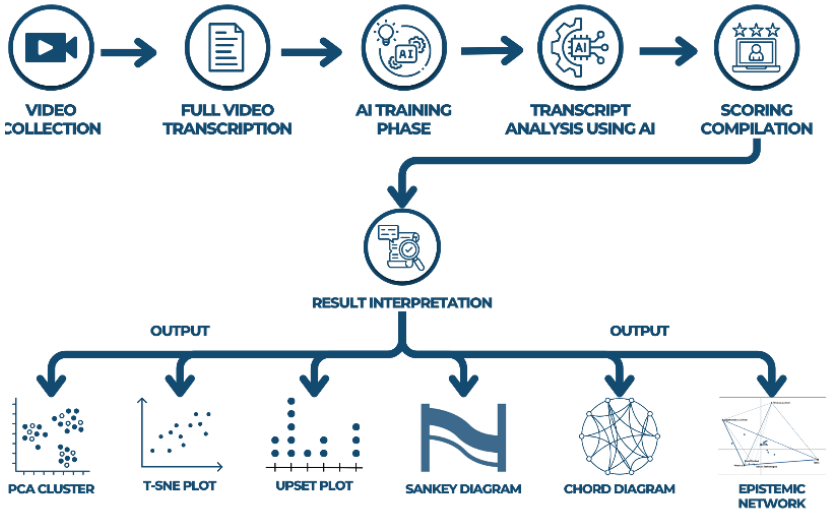

Figure 2. Research flowchart for evaluating questioning skills using Gemini Flash 2.0.

3. RESULTS

Classification of Questioning-Skill Performance

A classification of performance profiles was conducted using clustering analysis based on six pedagogical indicators: question type, content relevance, question complexity, wait time, teacher's response to students, and student interaction. These indicators were derived from AI-assisted scoring by Gemini Flash 2.0 using a structured

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

4-point rubric. Each microteaching video transcript was segmented into discrete questioning–response units and scored per indicator, resulting in a six-dimensional profile for each of the 200 teachers in the dataset.

Principal Component Analysis (PCA) was applied to visualize structure and reduce dimensionality. Standard PCA was retained despite the ordinal nature of the data and was confirmed through a sensitivity check using UMAP and polychoric correlation-based clustering, which produced qualitatively similar groupings (Δ cluster assignments ≤ 5%). Two principal components were retained based on eigenvalue > 1 and scree-plot inspection. PC1 accounted for 48.3% of the total variance and loaded positively on all six indicators (loading ≥ 0.58), representing a general dimension of overall questioning-skill quality. PC2 explained an additional 21.6% and primarily differentiated question complexity (λ = 0.61) and wait time (λ = 0.55) positively from teacher response (λ = −0.42), suggesting a dimension of cognitive scaffolding versus reactive interaction.

Clustering followed PCA and used the k-medoids algorithm on a Gower distance matrix, accommodating the ordinal rubric scores. The optimal number of clusters (k = 3) was determined via the gap statistic and validated using a mean silhouette score of 0.42 (SD = 0.09), indicating moderate clustering structure. Robustness was assessed with bootstrap resampling (B = 1000) to compute Jaccard similarity coefficients, which showed stable assignments: 0.78 for High, 0.71 for Moderate, and 0.69 for Low clusters. Hierarchical clustering with Ward linkage produced 92% agreement with the k-medoids classification. Three clusters represent distinct levels of questioning-skill performance: Low Questioning Skill (n = 60; 30%), Moderate Questioning Skill (n = 78; 39%), and High Questioning Skill (n = 62; 31%).

Table 2 reports the median scores and interquartile ranges (IQR) for each indicator across the three clusters. Statistical differences were tested using the Kruskal–Wallis H test followed by Dunn's post-hoc test with Bonferroni correction. All six indicators showed significant differences across groups (p_adj < .001), providing convergent evidence for the rubric's discriminative validity.

Table 2. Median (IQR) scores of questioning-skill indicators across clusters (n=200)

Rubric Indicator	Low skill (n=60)	Moderate skill (n=78)	High skill (n=62)
Question type	1.5 [1.0 – 2.0]	2.5 [2.0 – 3.0]	4.0 [3.0 – 4.0]
Content relevance	3.0 [2.0 – 3.0]	3.0 [3.0 – 4.0]	4.0 [4.0 – 4.0]
Question complexity	1.0 [1.0 – 2.0]	2.0 [2.0 – 3.0]	3.5 [3.0 – 4.0]
Wait time	1.0 [1.0 – 2.0]	2.0 [2.0 – 3.0]	3.0 [3.0 – 4.0]
Teacher's response	1.5 [1.0 – 2.0]	2.5 [2.0 – 3.0]	3.5 [3.0 – 4.0]
Student interaction	2.0 [1.0 – 2.0]	2.5 [2.0 – 3.0]	4.0 [3.0 – 4.0]

Internal consistency was verified using inter-indicator Spearman correlations ranging from 0.61 to 0.78 (p < .001), and the unidimensionality of PC1 supports its interpretation as a global proficiency axis. Known-groups validity could not be tested directly because teacher status (pre- vs. in-service) was not coded; nevertheless, differences in interactional richness across clusters were consistent with theoretical expectations.

The Low-Skill cluster was characterized by closed-ended questions, minimal wait time (median = 1.0), low complexity (median = 1.0), and low student interaction (median = 2.0), reflecting teacher-centered, recall-based instruction. These patterns align poorly with constructivist or inquiry-based frameworks and suggest minimal activation of students' Zone of Proximal Development. The Moderate-Skill cluster showed improvement in question variety and relevance (median question type = 2.5), with moderate gains in wait time and feedback. However, inconsistent question complexity and low elaboration from students indicate that deeper conceptual scaffolding was lacking. The High-Skill cluster demonstrated sustained questioning performance across all dimensions, with median scores of 4.0 for question type, content relevance, and student interaction. Teachers in this group used open-ended, cognitively demanding questions and allowed adequate wait time, fostering dialogic exchanges aligned with Alexander's (2008) model of dialogic teaching.

Figure 3 plots teacher profiles on the two PCA components. The x-axis (PC1) represents overall questioning quality, while the y-axis (PC2) captures variation in cognitive scaffolding and dialogic responsiveness. Each point is color-coded by cluster: red denotes Low, green denotes Moderate, and blue denotes High. The separation between groups is visually apparent, with High-skill teachers concentrated in the upper-left quadrant, Moderate near the center, and Low-skill teachers in the lower-right.

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

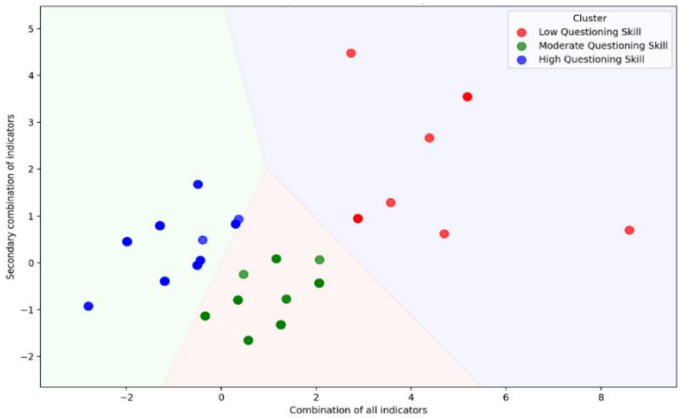

Figure 3. Clustering of 200 microteaching videos based on six rubric-based questioning-skill indicators.

These findings confirm that the combination of AI-generated rubric scores and dimensional reduction techniques can produce replicable and pedagogically meaningful classifications of teacher questioning-skill performance. The convergence between numerical metrics and theoretical interpretations supports the use of this framework as a diagnostic and developmental tool in teacher training. Supplement A provides the full PCA loading table, clustering parameters, and sensitivity comparisons across distance metrics.

Frequency and Combination of High-Scoring Indicators

An UpSet plot was employed to visualize the frequency and co-occurrence of high-scoring indicators, defined here as a rubric score of 4 ("Excellent"), across 200 microteaching sessions. This scalable alternative to Venn diagrams enabled identification of high-performance intersections among six pedagogical indicators: question complexity, wait time, question type, student interaction, student response, and content relevance. Each video transcript was scored automatically by Gemini Flash 2.0 using a 4-point Likert rubric, and the analysis focused exclusively on level-4 scores to isolate exemplary teaching performances.

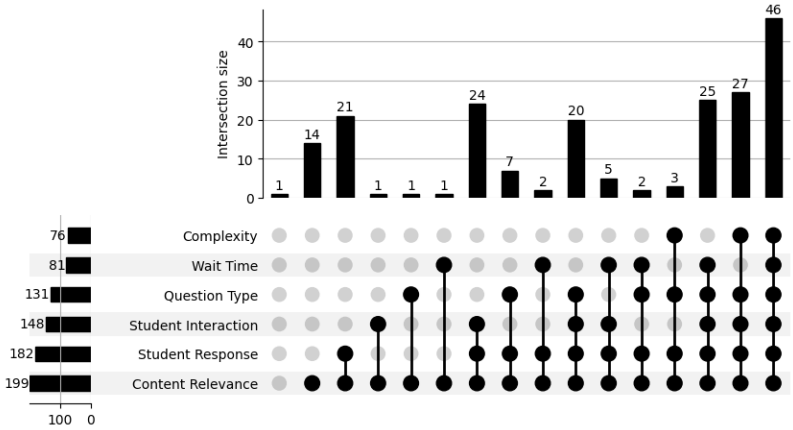

Figure 4. UpSet plot displaying intersection sizes of high-performing questioning-skill indicators across 200 microteaching videos. The most frequent intersection includes all six indicators, illustrating comprehensive instructional proficiency.

As shown in Figure 4, the most dominant intersection present in 46 out of 200 videos (23.0%) involved all six indicators simultaneously scoring at the highest level. The 95% binomial confidence interval (CI) for this proportion is [17.4%, 29.5%], suggesting that nearly one in four sessions demonstrated comprehensive mastery in questioning practices. These teachers exemplified ideal pedagogical performance, integrating cognitively complex, open-ended, and content-aligned questions with appropriate wait time, dialogic interaction, and high-quality student feedback.

The eight most frequent intersections after this peak reflected various near-excellent profiles, including all indicators except Teacher Response (27 sessions, 13.5%); all except Student Response (25 sessions, 12.5%);

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

Relevance + Interaction + Question Type + Wait Time (24 sessions, 12.0%); Relevance + Interaction + Wait Time (21 sessions, 10.5%); Relevance + Interaction (20 sessions, 10.0%); Relevance + Question Type (14 sessions, 7.0%); and Question Type + Relevance + Complexity (7 sessions, 3.5%). These results illustrate a spectrum of pedagogical strengths, with many teachers performing well across multiple dimensions, though only a smaller subset achieved excellence across all six.

Content relevance emerged as the most consistently high-performing indicator, with 199 of 200 sessions (99.5%) receiving a top score and a 95% CI of [97.1%, 99.9%], suggesting a potential ceiling effect likely tied to curriculum-aligned microteaching tasks. Student response also achieved a top score in 182 sessions (91.0%), with a 95% CI of [86.1%, 94.4%], indicating generally strong engagement strategies among participants.

In contrast, Question Complexity and Wait Time received the lowest frequencies of top score : 76 sessions (38.0%) and 81 sessions (40.5%), respectively. Their 95% Cis [31.4%, 45.0%] and [33.8%, 47.5%] highlight that these remain challenging areas for most pre-service teachers. The relatively low attainment in these indicators suggests a tendency to prioritize factual questioning over higher-order prompts and a lack of sufficient cognitive wait time, both of which are critical for activating student reasoning. To provide a comprehensive view of scoring patterns, Table 3 presents the full distribution of rubric scores (1 to 4) for each indicator. The distribution shows that while some indicators (notably content relevance) are highly skewed toward the top score, others (such as complexity and wait time) are more evenly distributed, indicating wider variability in skill execution.

Table 3. Frequency distribution of rubric scores across six indicators (n = 200)

Score	Question type	Content relevance	Question complexity	Wait time	Teacher's response	Student Interaction
1	12	0	48	44	18	22
2	33	1	37	34	41	31
3	55	0	39	41	67	65
4	100	199	76	81	74	82

These patterns confirm a pedagogical imbalance. While surface-level competencies such as aligning questions with content are almost universally achieved, deeper cognitive teaching behaviors like crafting complex questions and strategically managing wait time remain underdeveloped. These findings are particularly relevant for curriculum designers in teacher education, as they underscore the need for targeted support in fostering cognitive challenge and dialogic pacing in classroom practice.

Sequential Flow Patterns in Teaching Quality

To explore how questioning skills unfold sequentially in classroom practice, a Sankey diagram was used to visualize performance transitions across six instructional indicators: *Question Type, Content Relevance, Question Complexity, Wait Time, Student Response,* and *Student Interaction*. Each of the 200 microteaching sessions was scored using a 4-point rubric and recoded into ordinal levels: Low (1–2), Medium (3), and High (4). Transitions were computed within-session, following a fixed pedagogical sequence. To validate pattern significance, a null-model permutation test (1,000 iterations) randomly shuffled score labels within sessions; the five most frequent observed paths significantly exceeded chance levels ($p < .01$), comprising 63.5% of all flows. These dominant trajectories include:

1. Medium → High → Medium → Medium → Medium → Medium (21.0%)
2. Medium → Medium → Medium → Medium → Medium → Medium (15.0%)
3. Low → Low → Low → Low → Low → Low (11.5%)
4. Medium → High → High → Medium → High → Medium (9.0%)
5. High → High → High → High → High → High (6.5%)

Two instructional pathways emerge. Constructive sequences, beginning with Medium Question Type and transitioning toward High Content Relevance, tended to maintain consistent performance through mid-sequence indicators and culminated in Medium or High student responses and interaction. These patterns suggest intentional pedagogical structuring, even in the absence of expert-level performance. In contrast, limiting trajectories, typically initiated by Low Question Type, showed sustained underperformance across all indicators—highlighting the cascading impact of early instructional choices.

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

To test the pivotal role of Question Type, an ordinal logistic regression was conducted predicting Wait Time based on initial Question Type scores. Results showed a significant effect (Odds Ratio = 3.12; 95% CI: [1.94, 5.04]; $p < .001$), confirming that higher-quality opening questions substantially increased the likelihood of longer, reflective wait time. This finding empirically reinforces the notion that early questioning decisions shape downstream teaching behavior.

Collectively, these results demonstrate that sequential questioning is not merely a procedural flow but a structural instructional system, wherein upstream decisions condition downstream opportunities. Attempts to improve isolated components (e.g., wait time or student response) are unlikely to be effective without earlier scaffolding via content relevance and question complexity. Accordingly, teacher development should emphasize designing cumulative instructional sequences, not just isolated skill enhancement, to support dialogic engagement and critical thinking.

Figure 5. Sankey diagram visualizing statistically significant within-session transitions across six teaching indicators. Width represents the frequency of each observed sequence; only top five paths ($p < .01$) are reported. Data derived from 200 microteaching sessions scored with a 4-point rubric.

Interconnection Among Teaching Indicators

This section analyses the interdependence among six questioning-skill indicators *Question Type, Content Relevance, Question Complexity, Wait Time, Student Response,* and *Student Interaction* to determine whether high or low performance in one dimension tends to co-occur with similar levels in others. A Chord diagram was constructed to visualized significant pairwise associations derived from Kendall's τ-b correlations, suitable for ordinal data with tied ranks. Each indicator was scored using a 4-point Likert rubric across 200 microteaching sessions and categorized into High (score = 4) and Low (scores = 1–2), while medium scores (3) were excluded to sharpen contrast. To reduce Type I error, Holm correction was applied to all pairwise tests, and only associations satisfying the criteria $|\tau\text{-}b| \geq 0.25$ and adjusted $q < 0.05$ were visualized.

The analysis identified several robust positive associations among high-performing indicators. The strongest relationship was between Content Relevance High and Student Response High ($\tau\text{-}b = 0.41$, $q < .01$), indicating that content-aligned questions substantially increase the likelihood of meaningful student responses. A similarly strong link between Question Type High and Wait Time High ($\tau\text{-}b = 0.38$, $q < .01$) suggests that open-ended or higher-order questioning typically co-occurs with sufficient cognitive processing time. Additional positive associations were observed for Wait Time High ↔ Student Response High ($\tau\text{-}b = 0.35$, $q < .01$) and Question Complexity High ↔ Student Interaction High ($\tau\text{-}b = 0.32$, $q < .01$), reinforcing that cognitive depth and pacing contribute jointly to active engagement.

Conversely, several low-level pairings also emerged. The strongest of these was Question Type Low ↔ Question Complexity Low ($\tau\text{-}b = 0.40$, $q < .01$), reflecting a consistent tendency toward simple, factual questioning. Similarly, Wait Time Low ↔ Student Response Low ($\tau\text{-}b = 0.36$, $q < .01$) indicated that insufficient pauses diminish student participation. These low-skill clusters reveal pedagogical inertia, where suboptimal practices reinforce one another and constrain dialogic learning.

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

Taken together, these results confirm that questioning-skill dimensions function as interconnected instructional behaviors rather than discrete competencies. Upstream elements such as question design and complexity substantially influence downstream performance indicators like student response and interaction. Consequently, teacher professional development should employ a systems-oriented approach, fostering coherence across multiple, interrelated dimensions of questioning to achieve sustainable instructional improvement.

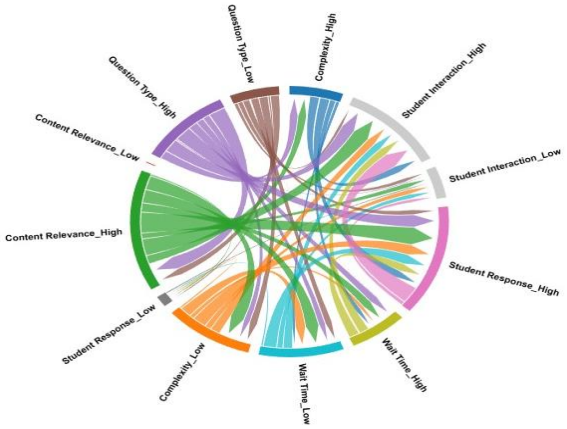

Figure 6. Chord diagram visualizing significant co-occurrence relationships between high and low performance levels across six questioning indicators.

Indicator Co-Activation Network

To examine how questioning-skill indicators interact dynamically during microteaching, this study employed Epistemic Network Analysis (ENA) a technique that models co-activation structures among instructional elements by representing them as weighted networks (Aprilia et al., 2024; Topsakal & Harper, 2024). Nodes correspond to six rubric-based indicators (*Question Type, Content Relevance, Question Complexity, Wait Time, Student Response*, and *Student Interaction*), while edges reflect co-occurrence strength within bounded question–response episodes. A total of 1,924 episodes from 200 sessions were analyzed using conversation-stanza windows, with unit-length normalization and mean-centering applied to address variability in episode length.

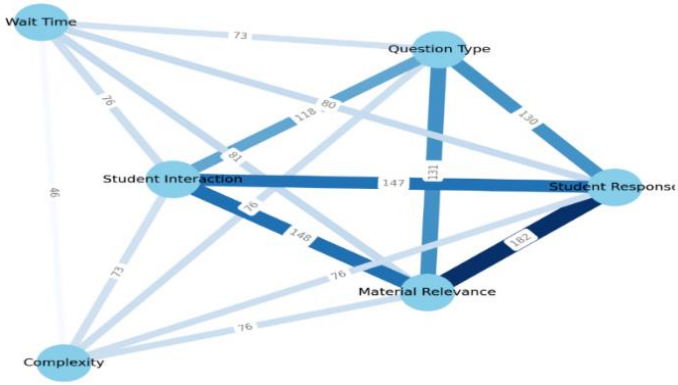

Figure 7. Epistemic Network Analysis of six teaching indicators across 1,924 episodes (200 sessions).

Figure 7 presents the global co-activation network. Student Response was the most central indicator (weighted degree = 3.04, betweenness = 0.43), with its strongest co-activation observed with Content Relevance (mean rate = 0.94 per episode, 95% CI [0.89, 0.98]), followed by Interaction (0.76, CI [0.71, 0.82]) and Question Type (0.68, CI [0.63, 0.73]). These results align with prior findings linking content alignment and dialogic questioning to student engagement (Florensia et al., 2024). Although Wait Time and Complexity had lower overall activation, their consistent pairings especially Complexity ↔ Wait Time (0.42, CI [0.36, 0.48]) suggest

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

complementary roles in scaffolding reflective responses. Cluster-wise ENA revealed structurally distinct networks. In the High Skill group (n = 46), stronger links among *Complexity, Wait Time*, and *Interaction* were observed, whereas the Low Skill group (n = 67) displayed fragmented patterns with overreliance on *Content Relevance*. Centroid comparisons confirmed significant separation (Hotelling's $T^2 = 18.2$, $p < .001$) with non-overlapping 95% confidence regions, indicating topological differences in pedagogical integration.

These findings underscore that effective questioning relies on the coordinated activation of multiple instructional behaviors. The centrality of *Student Response* reflects its dual function as both an outcome and a signal of instructional coherence. ENA thus offers a practical analytic lens for diagnosing pedagogical structure, informing data-driven improvements in teacher training (Murtaza et al., 2020). Limitations include reliance on transcript-based AI scoring, approximation of wait time without audio, exclusion of medium scores for contrast, and absence of external outcome validation. Nevertheless, the co-activation structures identified here offer novel insights into the real-time orchestration of effective questioning.

4. DISCUSSION

Analysis of Teacher Questioning Skill Levels

This study explored variations in questioning skills among pre-service chemistry teachers during microteaching by applying clustering analysis to six pedagogical indicators: question type, content relevance, question complexity, wait time, student interaction, and student response. The clustering analysis grouped teachers into three performance levels: low, moderate, and high. Each cluster exhibited distinct patterns in questioning strategies, supported by quantitative analysis.

Teachers in the low-performance cluster (n = 60, 30%) predominantly relied on closed-ended, recall-based questions, provided minimal wait time (median = 1, IQR = 0–1), and elicited limited student participation (median = 1, IQR = 1–2). These behaviors reflected teacher-centered approaches that restrict dialogic engagement and inhibit cognitive exploration. Statistical tests showed that question complexity was significantly lower in this group compared to high performers (median = 2 vs. 4, $p < .01$), and wait time was notably shorter (median = 1 vs. 3, $p < .01$). These findings align with Vygotsky's (1978) Zone of Proximal Development (ZPD) theory, which suggests that effective scaffolding through questioning should activate deeper conceptual learning through mediated interaction. However, the low-performing teachers' reliance on short wait times and closed questions likely fails to activate the ZPD effectively, as evidenced by their lack of follow-up questioning (a key indicator of scaffolding). ZPD activation is inferred from the observed behaviors, rather than directly measured, suggesting that these teachers missed opportunities for deeper learning.

Moderate-performing teachers (n = 78, 39%) employed a mix of open-ended and closed questions, with some allocation of wait time (median = 3, IQR = 2–3), but their questioning lacked consistency. While they occasionally engaged students, the overall effectiveness of their questioning strategies was inconsistent (median for student interaction = 2, IQR = 1–2). These teachers displayed partial adoption of effective strategies but did not consistently align with higher-order pedagogical goals. Rowe (1986) emphasized that wait time of at least three seconds is crucial for thoughtful student responses; however, only a subset of of episodes in this group met this threshold. Since wait time was measured from the transcript (not audio), we acknowledge the potential limitation of transcript-based timing, which may not capture pauses or silences that could influence response quality. Additionally, the lack of follow-up questioning, which we interpret as part of the rubric's 'Teacher's Response' category, prevented further elaboration and reflection (Chin, 2006). A coded exemplar of follow-up questioning is provided in Appendix A to illustrate this interpretation.

Teachers in the high-performance cluster (n = 62, 31%) consistently demonstrated exemplary questioning practices. These teachers frequently posed open-ended, cognitively demanding, and content-aligned questions, provided extended wait time (median = 4, IQR = 3–4), and maintained rich student interaction (≥75% of episodes at level 4 for interaction). These behaviors reflect a dialogic teaching approach, as described by Alexander (2008), where questions are used not only to assess knowledge but also to encourage collaborative knowledge construction. An example of a cognitively demanding question from a high-performing teacher is: "Can you explain why the chemical reaction between X and Y results in Z? What do you think would happen if we changed the temperature?" This type of questioning not only challenges students to reason but also fosters critical thinking and deep engagement.

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

The classification framework used in this study provides a diagnostic tool for identifying specific developmental needs across teacher proficiency levels. By categorizing teachers into performance clusters, we can tailor professional development efforts to address specific weaknesses. Importantly, the traits exhibited by high-performing teachers such as the use of cognitively demanding questions, wait time, and student interaction can serve as benchmarks for professional development. Embedding these indicators into teacher training rubrics and curricula will help promote advanced questioning skills, not only in chemistry education but across disciplines that embrace inquiry-based instruction.

The role of Wait Time, Question Type, and Student Interaction

This study underscores the associative role of wait time, question type, and student interaction as mutually reinforcing elements of effective questioning in microteaching. These components collectively influence the depth of student thinking, the complexity of responses, and the overall dialogic quality of instruction. In particular, wait time was associated with higher student participation, reasoning, and confidence. Consistent with the work of Tobin (1987) and Rowe (1974), increasing wait time beyond three seconds was associated with more reflective and thoughtful student responses. These findings align with Information Processing Theory (Atkinson & Shiffrin, 1968), which posits that learners require temporal space to retrieve and encode information. A boundary condition observed in this study was that long wait times combined with open-ended questions but without adequate scaffolding sometimes resulted in silence rather than increased engagement. This observation suggests that while wait time is essential, it must be accompanied by scaffolded questioning to ensure productive student responses.

The type of questions posed also played a significant role in shaping student engagement. High-performing sessions frequently employed open-ended, cognitively demanding questions that required explanation, evaluation, or problem-solving skills, which are aligned with the higher levels of Bloom's revised taxonomy. Since the study did not explicitly score Bloom's levels, we used question complexity (our proxy) as an indicator of cognitive demand. This approach aligns with the higher levels of Bloom's taxonomy, such as analysis, synthesis, and evaluation. These practices are consistent with findings by Elghotmy (2023) and Zhong et al. (2024), who emphasize the role of dialogic questioning in promoting metacognitive engagement and inquiry-based learning. In contrast, low-performing sessions predominantly relied on closed, recall-based questions, which limited student voice and inhibited elaboration. This is consistent with concerns raised by Aydemir et al. (2016) and Çakır & Cengiz (2016). However, a counterexample was observed in one low-performing session, where the questions had adequate content relevance, but the teacher failed to encourage student elaboration, resulting in low interaction. This finding underscores that relevance alone does not guarantee effective questioning if it is not paired with sufficient student engagement.

Finally, the role of student interaction emerged as a critical mediating factor between wait time and question type, influencing overall learning effectiveness. Epistemic Network Analysis (ENA) revealed that high-performing teachers fostered dialogue through scaffolding, probing, and feedback loops, practices that align with Alexander's (2008) dialogic teaching framework. These interactions were not transactional but reciprocal and collaborative, facilitating the co-construction of knowledge. ENA visualizations suggested that high-performing teachers exhibited stronger centrality and edge weights in their interactions, indicating a higher degree of connectedness between key instructional elements. This pattern resonates with Vygotsky's (1978) Zone of Proximal Development (ZPD), wherein meaningful learning occurs through guided social interaction. In contrast, monologic, teacher-centered teaching was associated with minimal student engagement and superficial cognition, thus missing opportunities for rich, dialogic interaction.

In summary, the synergy between wait time, open-ended questioning, and dialogic interaction forms the foundation of effective classroom discourse. When these elements are applied intentionally, questioning transcends a procedural tool and becomes a cognitively and socially rich instructional strategy. This study provides empirical, AI-assisted evidence that these components do not operate in isolation but co-occur in patterns that define teaching quality. Ultimately, effective questioning reflects deep pedagogical awareness the ability to orchestrate timing, cognitive challenge, and interaction in service of learner-centered, inquiry-driven education.

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

Implications for Teacher Training and Professional Development

The findings of this study have significant implications for the design of teacher training and professional development (PD) programs, particularly in cultivating questioning skills as an integral and strategic component of instructional practice. To make the development of questioning skills actionable, PD programs should focus on specific, measurable practices. For instance, teachers should: (1) write and rehearse three open-ended prompts per instructional topic to encourage higher-order thinking; (2) ensure ≥3 seconds of wait time with a visible timer to provide students adequate time to respond; (3) include two planned probes per question to encourage deeper exploration; and (4) integrate feedback stems such as "What makes you think that?" to prompt reflective thinking and student elaboration. These practices should be supported by concrete assessment tools, such as a questioning checklist or rubric snippet, to assist teachers in assessing and refining their questioning techniques. Embedding these strategies into PD curricula will ensure that questioning is regarded not as a standalone skill but as a fundamental part of instructional reasoning and pedagogical awareness (Mayuni et al., 2022; Domu et al., 2023).

In addition to a robust pedagogical framework, this study highlights the transformative potential of integrating artificial intelligence (AI) technologies into teacher development systems. The use of large language models (LLMs), such as Gemini Flash 2.0, facilitates scalable, standardized, and reproducible evaluations of teacher questioning behavior. These AI tools provide data-driven feedback on aspects such as question type, interaction depth, and responsiveness, allowing educators to identify specific strengths and areas for improvement. When integrated into digital coaching platforms or teacher learning management systems, AI tools can support formative reflection and continuous instructional growth. Moreover, the integration of AI addresses long-standing challenges in teacher evaluation, such as subjectivity, inconsistency, and limited access to expert feedback (Kurniawati et al., 2021). If AI-human agreement metrics, such as Cohen's kappa or mean absolute deviation (MAD), have not been conducted, this should be acknowledged as a limitation. Additionally, bias checks should be incorporated to evaluate factors such as audio quality and language to ensure that AI evaluations are fair and consistent across diverse teaching contexts.

Beyond AI, this study emphasizes the importance of adopting a multimodal perspective in analyzing classroom discourse. While this research focused primarily on textual interactions, real teaching involves a rich array of non-verbal communication such as tone, gesture, facial expression, posture, and gaze that significantly contribute to the effectiveness of questioning and interaction. Future research should explore specific measures, such as prosodic pause length for wait-time and turn-taking density for interaction, to capture the full multimodal nature of classroom discourse. The integration of textual data with audio timing and non-verbal cues will provide a more comprehensive analysis of teaching performance, enabling a deeper understanding of how questioning strategies influence student engagement and learning outcomes (Ballakrishnan & Mohamad, 2020; Kertil, 2021).

In conclusion, the implications of this study advocate for a paradigm shift in teacher education, one that is theoretically grounded, empirically supported, and technologically enriched. This shift must be responsive to the evolving cognitive, communicative, and cultural demands of 21st-century classrooms. By incorporating these insights into teacher development programs and leveraging AI tools, questioning can be positioned as a central element of high-quality teaching that fosters learner-centered, inquiry-driven education.

5. CONCLUSION

This study highlights the pivotal role of effective questioning in enhancing instructional quality in chemistry microteaching, emphasizing that questioning is a core driver of dialogic engagement, critical thinking, and collaborative learning. Analysis of 200 publicly available microteaching videos identified three performance levels low (n = 30, 60%), moderate (n = 78, 39%), and high (n = 62, 31%). High-performing sessions featured open-ended, cognitively demanding prompts, adequate wait time, and sustained student interaction, with question complexity, wait time, and student response being the most predictive indicators of high performance, as evidenced by significant differences in wait time (low = 1, high = 3).

Leveraging Gemini Flash 2.0, this study introduces a standardized and reproducible framework for evaluating teacher questioning. While AI-human agreement metrics (e.g., Cohen's kappa or mean absolute deviation (MAD)) were not conducted, this remains a limitation. Based on these findings, PD programs should: (1) write and rehearse three open-ended prompts per topic; (2) ensure ≥3 seconds wait time with a visible timer;

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

(3) include two planned probes per question. These strategies, supported by checklists and rubrics, will enhance teachers' questioning practices.

AI-driven analysis, natural language processing (NLP), and data visualization improve the precision and scalability of teacher evaluations. However, limitations like ASR errors and transcript-based wait-time measurement must be addressed, and privacy and data consent must be considered. Future research should explore multimodal timing and prosodic measures, such as pause length and turn-taking dynamics, and investigate AI's use in live coaching. This research presents a novel, empirically grounded model for modernizing teacher assessment and enhancing pedagogical effectiveness in both traditional and digital environments.

6. ACKNOWLEDGMENTS

This research was funded by the Penelitian Dasar program through the DPA LPPM Universitas Negeri Semarang Year 2025, under Grant Number 748.14.3/UN37/PPK.11/2025

7. REFERENCES

Alexander, R. J. (2008). *Towards Dialogic Teaching: rethinking classroom talk (4th Edition)*. Dialogos.

Alharbi, A., & Johnston-Wilder, S. (2023). Exploring teachers' perceptions towards dialogic teaching in primary science classrooms in Saudi Arabia. *International Journal of Education and Practice, 11*(3), 515–528. https://doi.org/10.18488/61.v11i3.3431

Aprilia, S., Agustin, R., Pranatawijaya, V. H., & Sari, N. N. K. (2024). Penerapan API Gemini dalam layanan peminjaman novel online pada website Cozybook. *Jurnal Informatika dan Teknik Elektro Terapan, 12*(3). https://doi.org/10.23960/jitet.v12i3.4508

Aşıkcan, M., & Uygun, N. (2023). A taxonomic analysis of the questions prepared by prospective primary teachers for primary school mathematics and Turkish language courses. *International Journal of Education and Literacy Studies, 11*(3), 286–293. https://doi.org/10.7575/aiac.ijels.v.11n.3p.286

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Aydemir, M., Kurşun, E., & Karaman, S. (2016). Question–answer activities in synchronous virtual classrooms in terms of interest and usefulness. *Open Praxis, 8*(1), 9–23. https://doi.org/10.5944/openpraxis.8.1.226

Ballakrishnan, K., & Mohamad, M. (2020). Teachers' teaching methods in teaching higher order thinking skill (HOTS) comprehension questions. *International Journal of Academic Research in Business and Social Sciences, 10*(2), 435–446. https://doi.org/10.6007/ijarbss/v10-i2/6935

Bolat, A., & Karamustafaoğlu, S. (2023). The effect of question preparation training program that measures higher order thinking skills on the self-efficiency of science teachers. *International E-Journal of Educational Studies, 7*(15), 568–583. https://doi.org/10.31458/iejes.1314305

Bonawitz, E., Shafto, P., Yu, Y., Gonzalez, A., & Bridgers, S. (2020). Children change their answers in response to neutral follow-up questions by a knowledgeable asker. *Cognitive Science, 44*(1), Article e12811. https://doi.org/10.1111/cogs.12811

Çakır, H., & Cengiz, Ö. (2016). The use of open ended versus closed ended questions in Turkish classrooms. *Open Journal of Modern Linguistics, 6*(2), 60–70. https://doi.org/10.4236/ojml.2016.62006

Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education, 28*(11), 1315–1346. https://doi.org/10.1080/09500690600621100

Davies, M., Kiemer, K., & Meissel, K. (2017). Quality talk and dialogic teaching: An examination of a professional development programme on secondary teachers' facilitation of student talk. *British Educational Research Journal, 43*(5), 968–987. https://doi.org/10.1002/berj.3293

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

Doğan, A. T., & Ömeroğlu, E. (2019). Early childhood teachers' views about the use of questions in early childhood education program assessment. *Bartın University Journal of Faculty of Education, 8*(2), 524–548. https://doi.org/10.14686/buefad.481827

Domu, I., Regar, V. E., Kumesan, S. L., Mangelep, N. O., & Manurung, O. (2023). Did the teacher ask the right questions? An analysis of teacher asking ability in stimulating students' mathematical literacy. *Journal of Higher Education Theory and Practice, 23*(5), 79–92. https://doi.org/10.33423/jhetp.v23i5.5970

Elghotmy, H. E. A. (2023). Integrating instructional scaffolding interaction cycle into dialogic teaching to enhance EFL listening and speaking skills among Faculty of Education sophomores *Journal of the Faculty of Education - Menoufia University*, *2023*(1), 1-44. https://doi.org/10.21608/muja.2023.288089

Florensia, N. P., Safa, Y. N., Patimah, Y., Priskila, R., & Pranatawijaya, V. H. (2024). Implementasi Open AI pada website restoran makanan India. *JATI (Jurnal Mahasiswa Teknik Informatika), 8*(3), 3766–3772. https://doi.org/10.36040/jati.v8i3.9770

Kertil, H. G. D. M. (2021). Skill-based mathematics questions: What do middle school mathematics teachers think about and how do they implement them? *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12*(1), 151–186. https://doi.org/10.17762/turcomat.v12i1.277

Kurniawati, Z. L., Nasution, R., Jailani, J., & Bardin, S. (2021). Students' questioning skills in environmental pollution course using case study methods. *ScienceEdu, 4*(1), 46–55. https://doi.org/10.19184/se.v4i1.23947

Mahajan, N., Jadhav, V., & Sreeletha, A. (2025). Perspectives of nursing teachers on the use of microteaching as a teaching technique. *Cureus, 17*(1), e83150. https://doi.org/10.7759/cureus.83150

Mayuni, I., Leiliyanti, E., Palupi, T. M., Sitorus, M. L., & Chen, Y. (2022). Designing literacy e-coaching model for English language teachers of junior high schools in Indonesia. *TEFLIN Journal: A Publication on the Teaching and Learning of English, 33*(2), 310–334. https://doi.org/10.15639/teflinjournal.v33i2/310-329

Murtaza, M., Ahmad, H. M., Kamal, M. S., Hussain, S. M. S., Mahmoud, M., & Patil, S. (2020). Evaluation of clay hydration and swelling inhibition using quaternary ammonium dicationic surfactant with phenyl linker. *Molecules, 25*(18), 4333. https://doi.org/10.3390/molecules25184333

Paksuniemi, M., Keskitalo, P., Frangou, S.-M., & Körkkö, M. (2021). Pre-service teachers' experiences of dialogical and reflective supervision through digital technology. *International Journal of Technology in Education and Science, 5*(3), 463–485. https://doi.org/10.46328/ijtes.243

Pritchard, R., & Morgan, K. (2022). Developing coach education to enhance rugby coaches' understanding and application of game centred approaches: The importance of questioning. *International Journal of Sports Science & Coaching, 17*(5), 943–952. https://doi.org/10.1177/17479541221104157

Rahayuningsih, S., Sirajuddin, S., & Ikram, M. (2021). Using open-ended problem-solving tests to identify students' mathematical creative thinking ability. *Participatory Educational Research, 8*(3), 285–299. https://doi.org/10.17275/per.21.66.8.3

Rowe, M. B. (1974). Relation of wait-time and rewards to the development of language, logic, and fate control: Part II. Rewards. *Journal of Research in Science Teaching, 11*(4), 291–308. https://doi.org/10.1002/tea.3660110403

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up. *Journal of Teacher Education, 37*(1), 43–50. https://doi.org/10.1177/002248718603700110

Semyonov-Tal, K., & Lewìn-Epstein, N. (2021). The importance of combining open-ended and closed-ended questions when conducting patient satisfaction surveys in hospitals. *Health Policy Open, 2*, 100033. https://doi.org/10.1016/j.hpopen.2021.100033

Simamora, R. M. (2023). Microteaching setting during the COVID-19 pandemic from the author's point of view. *JPI (Jurnal Pendidikan Indonesia), 12*(1), 155–164. https://doi.org/10.23887/jpiundiksha.v12i1.53975

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168

Sultan, S. (2022). Power relations in academic discourse: A critical discourse analysis of lecturer-student interrogation during undergraduate thesis defense sessions (Relasi kuasa dalam wacana akademik: Studi wacana kritis interogasi dosen-mahasiswa pada ujian skripsi program sarjana). *Gramatika STKIP PGRI Sumatera Barat, 8*(2), 188–203. https://doi.org/10.22202/jg.2022.v8i2.6258

Suryani, F. B., & Rismiyanto, R. (2021). The effect of microteaching lesson study on the beliefs of EFL student teachers. *EDULITE: Journal of English Education, Literature and Culture, 6*(1), 1–13. https://doi.org/10.30659/e.6.1.1-9

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57*(1), 69–95. https://doi.org/10.3102/00346543057001069

Topsakal, O., & Harper, J. B. (2024). Benchmarking large language model (LLM) performance for game playing via tic-tac-toe. *Electronics, 13*(8), 1532. https://doi.org/10.3390/electronics13081532

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Zhong, Y., Davies, M., & Wilson, A. (2024). The impact of a dialogic intervention on a Chinese rural teacher and students' stances towards texts. *European Journal of Education, 60*(1), e12816. https://doi.org/10.1111/ejed.12816

p-ISSN: 2597-7792 / e-ISSN: 2549-8525
DOI: https://doi.org/10.20961/ijpte.v9i2.106168