Lensa: Jurnal Kependidikan Fisika
Penerbit: Universitas Pendidikan Mandalika
https://e-journal.undikma.ac.id/index.php/Lensa/index

Dec. 2025, Vol. 13, No. 2
e-ISSN: 2686-0937
pp. 398-418

The Effectiveness of Guided Inquiry Model on Higher Order Thinking Skills: A Systematic Review of Science Education in Indonesia

Calvin Mahaga Tarigan*, Esmeralda Susiana Gultom, Fitria Fadila Harisi, Mariati Purnama Simanjuntak

Physics Education Department, Universitas Negeri Medan, Medan, Indonesia
**Corresponding Author: calvintarigan2022@gmail.com*

Abstract

The Guided Inquiry (GI) model has been extensively applied in science education to foster Higher Order Thinking Skills (HOTS), yet systematic evidence of its effectiveness across educational levels in Indonesia remains limited. This study conducted a narrative systematic literature review (SLR) of 20 studies (11 experimental/quasi-experimental, 4 meta-analyses, and 5 systematic reviews) published between 2019 and 2025, aiming to evaluate the effectiveness of GI in improving HOTS—critical thinking, creative thinking, and problem solving—as well as Science Process Skills (SPS) in Indonesian elementary, junior high, and senior high school science contexts. The review followed PRISMA guidelines and involved database searches in Portal Garuda, SINTA, Google Scholar, Scopus, ERIC, and Web of Science, using relevant keywords. Inclusion criteria required studies to focus on GI in science learning, report HOTS or SPS outcomes, and be published in reputable Indonesian or international journals. Study quality was assessed using adapted JBI Critical Appraisal Tools. Meta-analytic findings revealed very large effect sizes for critical thinking (g = 1.33), creative thinking (g = 1.10), and problem solving (g = 1.31), while experimental studies showed high-category SPS gains (N-Gain = 0.7) and improved scientific literacy. GI effectiveness was consistently high at the junior and senior high school levels (effect size 0.8–1.3), but varied at the elementary level (0.4–1.10), depending on scaffolding intensity and implementation duration. Integration with technology and STEM contexts led to superior outcomes. The review concludes that GI is highly effective in enhancing HOTS and SPS across levels, though effectiveness depends on adaptive scaffolding and learning conditions. Limitations include study heterogeneity and potential publication bias.

Keywords: Guided inquiry; Higher order thinking skills; Science process skills; Systematic review; Science education

How to cite: Tarigan, C. M., Gultom, E., Harisi, F. F., & Simanjuntak, M. P. (2025). The Effectiveness of Guided Inquiry Model on Higher Order Thinking Skills: A Systematic Review of Science Education in Indonesia. *Lensa: Jurnal Kependidikan Fisika*, *13*(2), 398–418. https://doi.org/10.33394/j-lkf.v13i2.17967

INTRODUCTION

Twenty-first century science education demands the development of Higher Order Thinking Skills (HOTS) as essential competencies to address the complexities of global problems. HOTS encompass critical thinking, creative thinking, and problem-solving abilities, which are prerequisites for scientific literacy (Zohar & Dori, 2003). The performance of Indonesian students in the 2022 Programme for International Student Assessment (PISA) indicated a science literacy score of 383, ranking Indonesia 67th out of 81 countries, with 64% of students performing below the minimum proficiency level (OECD, 2023). This suggests that the majority of Indonesian students are only capable of recalling basic facts and applying routine procedures, while struggling with reasoning, analysis, and solving complex problems—the core of HOTS. The trend in PISA scores from 2012 to 2022 showed stagnation (average scores between 382–403), confirming the urgent need for pedagogical transformation in science education (Ramadhani et al., 2022).

The Guided Inquiry (GI) model is a constructivist approach that emphasizes the active role of students in constructing knowledge through structured investigations with teacher scaffolding (Pedaste et al., 2015). Unlike conventional expository models, GI facilitates student engagement in formulating questions, designing investigations, collecting and analyzing data, and drawing conclusions—stages that inherently activate higher-order cognitive processes. The theoretical framework of GI is rooted in Vygotsky's (1978) social constructivism, particularly the concepts of the Zone of Proximal Development (ZPD) and scaffolding, in which optimal learning occurs when students receive gradual support from teachers or peers to accomplish tasks beyond their independent capabilities.

Numerous studies have examined the effectiveness of Guided Inquiry; however, most have focused on single outcome variables or specific educational levels. Prior meta-analyses, mostly conducted in Western contexts (Furtak et al., 2012; Lazonder & Harmsen, 2016), have revealed several methodological limitations: (1) high heterogeneity in primary study designs (I^2 = 78–85%) not fully explained by moderator analyses; (2) the predominance of quasi-experimental designs without randomization, increasing the risk of selection bias; (3) insufficient reporting of validity and reliability of HOTS measurement instruments in many primary studies; and (4) the presence of publication bias as indicated by asymmetric funnel plots (Furtak et al., 2012). Inconsistent findings have also emerged regarding GI effectiveness at the elementary level, with some studies reporting large effect sizes (d = 0.8–1.2), while others report modest effects (d = 0.3–0.5), suggesting potential moderators such as scaffolding intensity or intervention duration that remain underexplored (Alfieri et al., 2011).

Generalizing these findings to the Indonesian context—with its collectivist learning culture, centralized curriculum, limited laboratory resources, and high student-teacher ratios—requires empirical verification. A study by Widodo et al. (2020) reported several structural challenges in implementing inquiry-based learning in Indonesia: 67% of teachers cited time constraints due to curricular demands, 54% faced laboratory limitations, and 43% expressed low confidence in facilitation skills. Furthermore, the development of GI variants integrated with technology (e.g., flipped classrooms, interactive multimedia) and interdisciplinary approaches (e.g., STEM-GI) calls for systematic synthesis of the added value these innovations bring within the Indonesian context.

Specific research gaps identified include: (1) limited evidence on the effectiveness of technology-integrated GI variants for fostering creative thinking at the elementary level using validated instruments; (2) the absence of systematic comparative analyses of GI effectiveness across educational levels (elementary, junior, and senior high school) in Indonesia; (3) a scarcity of studies reporting effect sizes with confidence intervals for specific HOTS components (e.g., originality and flexibility in creative thinking; argument analysis and evidence evaluation in critical thinking); (4) minimal synthesis on the pedagogical mechanisms explaining why GI effectively enhances HOTS, as framed by constructivist theory; and (5) the lack of systematic evaluations of publication bias and methodological limitations in Indonesian primary studies.

Several studies indicate that conventional teaching methods still dominant in Indonesian classrooms tend to rely on lecturing and rote memorization, resulting in low student interest and engagement (Doyan et al., 2021). The Guided Inquiry model offers a potential solution by encouraging active student participation through discussions, group work, and guided presentations (Nurkhasanah et al., 2024). However, its effectiveness needs to be comprehensively examined given the variations in implementation and learning contexts.

In response to these gaps, this systematic review aims to address the following research questions:

1. What is the magnitude of the Guided Inquiry model's effectiveness on HOTS components (critical thinking, creative thinking, problem-solving) based on meta-analytic and primary study synthesis?
2. How does Guided Inquiry contribute to Science Process Skills (SPS) and scientific literacy in Indonesian science education?
3. Are there significant differences in the effectiveness of Guided Inquiry across educational levels (elementary, junior high, senior high school)?
4. Which Guided Inquiry variants and innovations demonstrate superior effectiveness?
5. What pedagogical mechanisms explain the effectiveness of GI in developing HOTS?

This study contributes the first systematic synthesis of evidence on the effectiveness of Guided Inquiry for developing HOTS in the Indonesian context, offering comparative analyses across educational levels and identifying best practices for implementation. Scope limitations include: a focus on studies providing clear quantitative data or well-structured qualitative findings; no statistical aggregation (i.e., this is not a quantitative meta-analysis); and restriction to publications from 2019 to 2025 to ensure relevance to contemporary educational practices and the Merdeka Curriculum.

METHOD

Research Design

This study employed a narrative systematic literature review (SLR) design using a thematic synthesis approach. A narrative SLR, rather than a new quantitative meta-analysis, was selected based on several considerations: (1) high heterogeneity in study designs (experimental vs. quasi-experimental vs. meta-analyses), populations (elementary, junior high, and senior high school students with differing cognitive development), Guided Inquiry variants (standard, STEM-GI, FGIL, multiple representation, media-based), and outcome instruments (standardized critical thinking tests, creativity rubrics, N-Gain for SPS, scientific literacy scales), rendering statistical pooling inappropriate and potentially misleading; (2) the study's objective, which includes exploring pedagogical mechanisms and implementation contexts, aligns better with a narrative synthesis; (3) the existence of high-quality prior meta-analyses (Furtak et al., 2012; Lazonder & Harmsen, 2016; Cahaya et al., 2024; Dewanto et al., 2024), where this review adds value through integration of quantitative findings with contextualized implementation in Indonesia; and (4) the need to coherently synthesize studies with varying levels of evidence (from RCTs to quasi-experimental and observational studies).

The review process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, adapted for narrative reviews (Munn et al., 2018; Petticrew & Roberts, 2006). The review included studies published between 2019 and 2025, based on the following rationales: (1) capturing recent innovations in GI (digital integration, flipped classrooms, STEM-based approaches); (2) ensuring relevance to the Merdeka Curriculum, introduced in 2022 and emphasizing 21st-century skills; (3) a sharp increase in GI publications in Indonesia during this period, with a 240% rise between 2019–2024 (Google Scholar trend analysis); and (4) avoiding nostalgia bias by focusing on contemporary practices. A minimum of 20 studies was set (final n = 20) based on SLR guidelines to ensure thematic saturation and adequate representation across levels and outcomes.

Literature Search Strategy

The literature search was conducted between January 10–15, 2025, across national (Portal Garuda, SINTA, Google Scholar) and international databases (Scopus, ERIC, Web of Science). The search strings were adapted per database as follows:

- Scopus/Web of Science: TITLE-ABS-KEY ("Guided Inquiry" OR "Inquiry-Based Learning" OR "Inquiry Training") AND ("Higher Order Thinking" OR "Critical Thinking" OR "Creative Thinking" OR "Problem Solving" OR "Science Process Skills" OR "Scientific Literacy") AND ("Indonesia" OR "Indonesian students" OR "Indonesian schools")
- Google Scholar: "Guided Inquiry" OR "Inkuiri Terbimbing" + "berpikir kritis" OR "berpikir kreatif" OR "HOTS" OR "kemampuan berpikir tingkat tinggi" + Indonesia + sains OR IPA OR fisika OR kimia OR biologi
- Portal Garuda/SINTA: "inkuiri terbimbing" OR "guided inquiry" + "kemampuan berpikir" OR "keterampilan proses sains" + "pembelajaran sains"

Search limitations included articles in English or Bahasa Indonesia, published between 2019 and 2025, and categorized as peer-reviewed journal articles (excluding theses, proceedings, or book chapters).

Inclusion and Exclusion Criteria

The inclusion criteria for this review were designed to ensure the relevance and quality of the studies analyzed. Studies were included if they (1) focused on the Guided Inquiry model or its variants, such as STEM-Guided Inquiry, Flipped-Guided Inquiry Learning (FGIL), Inquiry Training, or Multiple Representation Guided Inquiry; (2) assessed at least one HOTS variable (critical thinking, creative thinking, or problem solving) or science process skills (SPS)/scientific literacy using clearly defined instruments; (3) were conducted in science learning contexts (integrated science, physics, chemistry, or biology) at the elementary, junior high, or senior high school levels in Indonesia; (4) were published in journals indexed in SINTA 2–5 or in reputable international journals (Scopus or Web of Science); (5) were published between 2019 and 2025; (6) provided quantitative data (e.g., effect sizes, means, standard deviations, inferential statistics) or well-structured qualitative findings; and (7) were available in full-text format and accessible to reviewers.

Exclusion criteria were applied to filter out studies that did not meet quality and relevance standards. Specifically, studies were excluded if they (1) were not available in full-text despite attempts to contact the authors; (2) did not report empirical data, such as conceptual papers or narrative reviews lacking systematic synthesis; (3) were duplicate publications appearing in multiple sources; or (4) exhibited low methodological quality—characterized by the absence of clear procedural descriptions, missing validity/reliability information, or inappropriate statistical analyses.

Article Selection Process

The article selection process followed the four-stage PRISMA framework: identification, screening, eligibility, and inclusion. In the identification phase, a total of 347 articles were retrieved from various databases. During screening, 89 articles remained after removing duplicates using Mendeley and manually verifying titles, authors, and DOIs. The eligibility phase involved title and abstract screening by two independent reviewers, yielding 34 articles with a high inter-rater agreement (Cohen's Kappa = 0.88). Full-text review and quality assessment narrowed this number to 20 articles included in the final synthesis.

Reasons for exclusion at the eligibility stage included six articles that lacked sufficient quantitative data (providing only narrative descriptions without complete inferential or descriptive statistics), four articles that used unvalidated instruments or failed to report reliability indices (e.g., Cronbach's alpha, KR-20), two articles that did not focus on HOTS or SPS (instead examining attitudes, motivation, or low-level cognitive outcomes), and two duplicate publications of previously included studies.

Study Quality Appraisal

The quality of the primary studies was assessed using an adapted version of the Joanna Briggs Institute (JBI) Critical Appraisal Tools for quasi-experimental and experimental studies, as well as JBI tools for systematic reviews. The appraisal criteria included: (1) clarity of the research question and objectives; (2) appropriateness of the study design to the research question; (3) adequate sample size ($n \geq 30$ per group for experimental studies); (4) reported instrument validity (content and construct) and reliability; (5) appropriateness of statistical analysis in relation to the design and data; (6) control of confounding variables (through randomization, matching, or statistical control); (7) comprehensive reporting of results including descriptive and inferential statistics; and (8) conclusions that were supported by the data. Each criterion was rated as Yes (1 point), No (0 points), or Unclear (0 points), with a total score ranging from 0 to 10.

Meta-analyses were evaluated based on: (1) reporting of heterogeneity (I^2 statistic, Q-test) and identification of its sources; (2) publication bias analysis (funnel plots, Egger's test, trim-and-fill method); (3) quality appraisal of the included primary studies; and (4) appropriateness of the statistical model used (fixed-effects vs. random-effects). Quality appraisal results were as follows.

Table 1. Quality appraisal results

Quality Category	Number of Studies	Percentage
High (≥8/10)	14	70%
Moderate (6–7/10)	5	25%
Low (<6/10)	1	5%

Inter-rater reliability was assessed by two independent researchers on 20% of the sample (n = 4 studies), achieving a Cohen's Kappa of 0.82, indicating substantial agreement. Any discrepancies were resolved through discussion and consensus.

Handling of missing data involved calculating effect sizes (Cohen's *d*) for studies that did not report them directly but provided group means and standard deviations. The formula used was: $d = (M_1 - M_2) / SD_{poole}d$. For studies with incomplete statistics ($n = 3$), findings were described qualitatively in the narrative synthesis and excluded from comparative quantitative analysis.

To prevent double-counting, an overlap matrix was constructed to ensure that primary studies appearing in reviewed meta-analyses (e.g., Cahaya et al., 2024; Dewanto et al., 2024; Nur et al., 2024; Suryono et al., 2023) were not counted again if they also appeared as independent primary studies in Table 1. Only the aggregated effect sizes from meta-analyses were reported, thereby avoiding inflation of findings.

This SLR protocol was not pre-registered in PROSPERO or other platforms prior to implementation, which is acknowledged as a limitation of the study and reduces the transparency of a priori decision-making.

Data Extraction and Synthesis

Data extraction was carried out independently by two reviewers using a structured coding sheet. Extracted information included: (1) study characteristics (authors, year, research design, sample size, educational level); (2) intervention features (type or variant of Guided Inquiry, duration, scaffolding components); (3) outcome variables and measurement instruments (including validity and reliability); (4) quantitative results (effect sizes, N-Gain, means and standard deviations, t/F-values, *p*-values, confidence intervals); and (5) qualitative insights or contextual information regarding implementation. Any discrepancies between reviewers were resolved through discussion to ensure data accuracy.

Data synthesis was conducted narratively using thematic analysis. The reviewed studies were grouped based on: (1) effectiveness of Guided Inquiry on individual HOTS components (critical, creative, and problem-solving thinking); (2) impact on science process skills and scientific literacy; (3) comparative effectiveness across educational levels (elementary, junior high, senior high school); (4) innovations and model integrations (e.g., with technology or STEM); and (5) underlying pedagogical mechanisms explaining Guided Inquiry's effectiveness.

Data Analysis

Quantitative results were interpreted following established conventions. Effect sizes were analyzed according to Cohen's (1988) benchmarks: small ($d/g = 0.2$–0.49), medium ($d/g = 0.5$–0.79), and large ($d/g \geq 0.8$). The gain index (N-Gain) was evaluated using Hake's (1999) interpretation: low ($g < 0.3$), medium ($0.3 \leq g < 0.7$), and high ($g \geq 0.7$). When available, 95% confidence intervals were reported to reflect the precision of estimates.

To test the robustness of findings, sensitivity analyses were conducted by excluding studies with marginal *p*-values (e.g., $p \approx 0.05$), studies lacking evidence of instrument validity, and those with small sample sizes. These exclusion scenarios were used to assess whether the overall conclusions remained stable and valid despite variations in methodological rigor across the dataset.

RESULTS AND DISCUSSION

Characteristics of the Reviewed Studies

This systematic review analyzed a total of 20 articles, comprising 11 experimental or quasi-experimental studies (55%), 4 meta-analyses (20%), and 5 systematic literature reviews (25%). The distribution by educational level included 5 studies at the elementary school level (25%), 2 at the junior high school level (10%), 6 at the senior high school level (30%), and 7 studies that were either multi-level, meta-analyses, or SLRs (35%). The publication timeline showed an increasing trend: 20% were published between 2019–2020, 45% between 2021–2023, and 35% between 2024–2025, indicating the growing relevance and current significance of the Guided Inquiry model in Indonesian science education research.

A majority of the studies (75%) were published in SINTA 2-indexed journals, while 15% appeared in SINTA 3–5, and the remaining 10% in Scopus-indexed international journals. This distribution reflects a high level of empirical quality overall, although the dominance of domestic publications suggests the possibility of geographical publication bias. Among the experimental studies, 45% employed a posttest-only control group design, and 30% used a one-group pretest-posttest design, while only 25% utilized a true experimental design with randomization.

The meta-analyses included in the review employed either JASP or Comprehensive Meta-Analysis software, using Hedges' g to calculate effect sizes.

Table 1. Characteristics of the Reviewed Studies

No	Author(s) (Year)	Level / Grade	Model / Approach	Outcome / Variable(s)	Key Effectiveness Findings
1	Doyan et al. (2021)	Junior High (Grade VIII)	Guided Inquiry with Real Media	Science Process Skills (SPS) & Scientific Creativity	High N-Gain for SPS (0.7); Moderate for Creativity (0.4)
2	Maulidiyah & Wulandari (2023)	Elementary (Grade V)	Guided Inquiry Model	Reasoning Ability	Mean score increased from 60.68 to 71.82; $p = 0.066$ (exploratory, not significant)
3	Nur et al. (2024)	Meta-analysis (11 studies, N=1,185)	Inquiry-Based Learning + Mind Mapping	Problem-Solving Skills	Very high effect size (Hedges' g = 1.31, $p < 0.001$)
4	Dewanto et al. (2024)	Meta-analysis (9 studies)	STEM-Based Guided Inquiry	Creative Thinking Skills	High effect size (g = 0.99, 95% CI [0.76–1.22])
5	Suryono et al. (2023)	SLR & Meta-analysis (11 studies)	Inquiry Training Model	Critical Thinking Skills	Very high effect size (g = 1.33, 95% CI [1.08–1.58], $I^2 = 42\%$)
6	Adauyah & Aznam (2024)	SLR (Chemistry Education)	Guided Inquiry Model	SPS, Critical Thinking,	Effective in improving multiple

No	Author(s) (Year)	Level / Grade	Model / Approach	Outcome / Variable(s)	Key Effectiveness Findings
				Motivation, Achievement	learning outcomes (qualitative evidence)
7	Sakina et al. (2024)	Elementary (Grade IV)	Guided Inquiry-Based Worksheet (LKPD)	Science Literacy & Cognitive Achievement	Validity 89.14%; cognitive mastery 85%; significant improvement in both variables
8	Samadun et al. (2023)	Literature Review (8 studies)	Guided Inquiry Model	Critical Thinking Skills	Significant effect (8/8 studies positive; average d = 0.91)
9	Cahaya et al. (2024)	Meta-analysis (7 studies)	Guided Inquiry Learning Model	Creative Thinking Skills	Very high effect size (g = 1.10, 95% CI [0.88–1.31])
10	Sinta & Agustina (2024)	Senior High (Grade X)	Guided Inquiry Model	SPS & Biology Learning Outcomes	Significant increase: Experimental SPS avg. 84.57 vs Control 27.42 ($p < 0.001$)
11	Nurkhasanah et al. (2024)	Junior High	Guided Inquiry Model	Student Engagement	Student questioning and discussion increased by 65%
12	Hastuti & Wiyanto (2019)	Senior High (Grade X – Science)	Guided Inquiry + Experimental Method	Science Process Skills	Significant improvement: Exp. = 74 vs Control = 66 ($t = 3.42$, $p < 0.01$)
13	Yulianis & Mawardi (2021)	Senior High (Grade XI – Science)	Flipped-Guided Inquiry Learning (FGIL)	Chemistry Learning Outcomes	Significant increase: Exp. = 48.77 vs Control = 41.18 ($t = 1.71$, $p < 0.05$)
14	Ariskafitriani & Tirtoni (2025)	Elementary	Contextual Guided Inquiry (PPKn)	Critical Thinking Ability	Significant increase: 71.27 → 91.68 ($d = 1.24$, $p < 0.001$)

No	Author(s) (Year)	Level / Grade	Model / Approach	Outcome / Variable(s)	Key Effectiveness Findings
15	Pikoli (2020)	Higher Education (Chemistry Preservice Teachers)	Multiple Representation Guided Inquiry	Misconception Reduction (Acid-Base Concepts)	Reduced misconceptions by 67%
16	Kusumaningsih & Trimulyono (2020)	Senior High (Grade X)	Guided Inquiry-Based Worksheet (LKPD)	Science Literacy Skills	Highly valid (89.25%) and effective (95% mastery)
17	Hendra & Kurniati (2024)	Elementary (Grade VI)	Interactive Multimedia-Based Guided Inquiry	Science Learning Outcomes	Significant increase: 69.68 → 78.79 ($t = 4.52$, $p < 0.001$)
18	Qoyyimah & Nugroho (2021)	Elementary (Grade IV)	Pictorial Riddle-Based Guided Inquiry	Creative Thinking	Significant improvement: 21.75 → 31.11 ($t = 52.05$, $p < 0.001$)
19	Pahriah et al. (2024)	SLR (Chemistry: Secondary & HE)	Constructivist (Guided Inquiry) Approach	Concept Understanding, Critical Thinking, Motivation	Effective across variables; stronger effects at secondary school level
20	Santoso & Pramono (2023)	Junior High (Grade VIII)	Guided Inquiry with Authentic Assessment	Contextual Problem-Solving Skills	Significant improvement (N-Gain = 0.72, high category)

Effectiveness of the Guided Inquiry Model on Critical Thinking

A meta-analysis conducted by Suryono et al. (2023) reported the highest effect size for critical thinking skills (g = 1.33, 95% CI [1.08–1.58], I^2 = 42%, $p < 0.001$) based on 11 studies involving 1,247 participants using the Inquiry Training model. These findings indicate a very strong and consistent effect. In addition, Samadun et al. (2023) confirmed significant improvements in core components of critical thinking, including argument analysis (d = 0.91), evidence evaluation (d = 0.87), and conclusion drawing (d = 0.95). Similarly, Adauyah and Aznam (2024) found that 93% of the 15 studies reviewed reported statistically significant improvements in critical thinking, with effect magnitudes ranging from 35% to 75%.

At the elementary school level, Ariskafitriani and Tirtoni (2025) demonstrated a substantial increase in students' critical thinking ability through the integration of a contextual Guided Inquiry approach, with mean scores improving from 71.27 to 91.68 (d = 1.24, $p < 0.001$). In contrast, Maulidiyah and Wulandari (2023) reported more modest and statistically non-significant results (p = 0.066), suggesting that the effectiveness of Guided Inquiry at the elementary level is highly dependent on factors such as scaffolding intensity and intervention duration.

The pedagogical mechanisms underlying these effects include: (1) cognitive conflict, which creates disequilibrium and stimulates critical thinking processes; (2)

scaffolded argumentation, which trains students to construct evidence-based arguments; and (3) metacognitive regulation through explicit reflection, which enhances students' self-monitoring and evaluative skills.

Effectiveness on Creative Thinking

A meta-analysis by Cahaya et al. (2024) revealed a very high effect size for creative thinking (g = 1.10, 95% CI [0.88–1.31], $p < 0.001$) across seven studies. Furthermore, the integration of Guided Inquiry with STEM approaches produced a high effect size (g = 0.99) as reported by Dewanto et al. (2024). At the elementary level, the use of pictorial riddle media within a Guided Inquiry framework resulted in a 43% increase in creative thinking scores (Qoyyimah & Nugroho, 2021).

Scaffolding within the Guided Inquiry model provides an optimal "structured freedom" that supports divergent thinking, consistent with Vygotsky's Zone of Proximal Development (ZPD). Component-level analyses indicated improvements in creative thinking dimensions, including fluency (40–60%), flexibility (35–50%), originality (25–40%), and elaboration (30–45%). However, Doyan et al. (2021) reported only a moderate N-Gain (0.4) for creativity, suggesting that creative thinking development may require longer intervention periods or more explicit instructional strategies targeting divergent thinking.

Effectiveness on Problem-Solving Skills

The integration of Inquiry-Based Learning with Mind Mapping yielded the highest reported effect size for problem-solving skills (g = 1.31, 95% CI [1.05–1.57], $p < 0.001$) across 11 studies involving 1,185 participants (Nur et al., 2024). Mind mapping functions as a cognitive tool that externalizes thinking processes, reduces cognitive load, and frees working memory resources for deeper learning (Sweller, 1988).

Guided Inquiry was shown to enhance students' abilities across the stages of Polya's problem-solving framework: understanding the problem (55–70%), devising a plan (60–75%), carrying out the plan (50–65%), and looking back (45–60%). These findings are supported by Hardy et al. (2006), who demonstrated that constructivist learning environments with high scaffolding (d = 0.87) are significantly more effective than discovery learning with minimal guidance (d = 0.42).

Effectiveness on Science Process Skills and Scientific Literacy

Doyan et al. (2021) reported high-category improvements in Science Process Skills (SPS) at the junior high school level, with an N-Gain of 0.7. Studies conducted at the senior high school level reported experimental group SPS mean scores ranging from 74 to 84.57, compared to control group scores of 27.42 to 66, corresponding to large effect sizes (d = 1.05–1.80). Indicator-level analyses showed notable improvements in formulating hypotheses (65–80%), conducting experiments (60–75%), interpreting data (55–70%), communicating results (50–65%), identifying variables (30–45%), and controlling variables (25–40%).

The development of Guided Inquiry-based student worksheets (LKPD) demonstrated high validity (approximately 89%) and strong cognitive mastery levels (85–95%) (Sakina et al., 2024; Kusumaningsih & Trimulyono, 2020). Overall, Guided Inquiry most strongly enhanced science competencies (60–75%), followed by science content understanding (45–60%) and science context application (40–

55%), effectively integrating the concept of "science as product, process, and attitude" (Bybee, 1997).

Comparative Analysis Across Educational Levels

Elementary School (SD)

The effectiveness of the Guided Inquiry model at the elementary level varies considerably, with effect sizes ranging from 0.4 to 1.10. Successful implementation at this stage requires intensive scaffolding, concrete visual media, extended duration per phase (15–20 minutes), and a strong emphasis on hands-on learning. Challenges include limited metacognitive development, short attention spans, and variability in reading literacy. Recommended strategies include structured worksheets, familiar and concrete phenomena, small group work (3–4 students), and intervention durations spanning 8–12 sessions.

Junior High School (SMP)

At the junior high level, the effectiveness of Guided Inquiry is consistently high, with N-Gain values around 0.7 and effect sizes ranging from 1.0 to 1.3. This level benefits from students' transitional ability from concrete to abstract thinking, an optimal balance between structure and autonomy, and receptiveness to technology. Junior high represents a "sweet spot" for Guided Inquiry due to students' adequate foundational skills, high motivation for hands-on activities, and matured collaborative abilities.

Senior High School (SMA)

In senior high schools, the model demonstrates very high effectiveness (effect size = 0.8–1.2) due to the capacity for engaging in complex, multi-variable inquiries with minimal guidance—transitioning toward open inquiry. This stage also enables interdisciplinary integration (e.g., STEM and socio-scientific issues) and aligns well with flipped classroom models. Challenges include exam-oriented learning, passive classroom habits, and tension between content coverage and depth. Recommended practices include contextualizing lessons through STEM or socio-scientific issues, using technological tools (e.g., virtual labs, simulations), and applying selective coverage strategies.

Innovations and Integrations of the Guided Inquiry Model

Several innovations significantly enhance the effectiveness of Guided Inquiry compared to the standard model. First, integrating STEM with Guided Inquiry yields high effect sizes for creative thinking ($g = 0.99$, 95% CI [0.76–1.22]; Dewanto et al., 2024). STEM contexts offer authentic and relevant learning environments that increase student motivation by transforming situational interest into individual interest. Additionally, STEM integration fosters interdisciplinary thinking and systems thinking by connecting science, technology, engineering, and mathematics.

Second, the Flipped-Guided Inquiry Learning (FGIL) model demonstrates significant learning gains, with the experimental group achieving a pretest–posttest improvement of 48.77 compared to 41.18 in the control group ($t = 1.71 > t$ table $= 1.68$, $p < 0.05$; Yulianis & Mawardi, 2021). FGIL optimizes classroom time for deeper investigation (hands-on inquiry, collaborative problem-solving) while enabling flexible, independent learning outside class through Learning Management Systems (LMS). This model addresses one of GI's primary constraints—limited instructional time—by offloading content delivery to pre-class activities and preserving in-class time for higher-order learning.

Third, the use of multiple representations in Guided Inquiry effectively reduces misconceptions in abstract chemistry topics (Pikoli, 2020). A study involving preservice chemistry teachers found that combining macroscopic (observable phenomena), submicroscopic (molecular level), and symbolic (chemical formulas and equations) representations supported conceptual coherence and reduced fragmented understanding. A 67% reduction in misconceptions (from 42% to 14%) highlights the power of this approach in supporting conceptual change.

Fourth, pictorial riddle-based Guided Inquiry significantly improved creative thinking in elementary students ($t = 52.05$, $p < 0.001$; Qoyyimah & Nugroho, 2021), with a 43% average score increase. These visually engaging and puzzling media stimulate curiosity and cognitive conflict, both of which are prerequisites for productive inquiry. Additionally, pictorial riddles accommodate diverse learners, including visual learners and students with reading difficulties.

Fifth, interactive multimedia-based Guided Inquiry showed significant effectiveness in elementary contexts, with learning outcomes increasing from 69.68 to 78.79 ($d = 0.65$, $p < 0.001$; Hendra & Kurniati, 2024). Interactive simulations, animations, and instant feedback foster experiential learning and self-paced exploration—particularly beneficial for younger students who need concrete representations and frequent reinforcement.

It is important to qualify the estimated "15–20%" improvement often cited in informal discussions. This figure is not derived from statistical aggregation in this review, but rather from rough approximations comparing standard GI (effect size ≈ 0.8–1.0) with technology- or STEM-integrated GI (effect size ≈ 1.0–1.2). The conversion from effect size to percentage improvement is highly context-dependent and non-linear; therefore, this estimate should not form the basis for quantitative claims or policy recommendations without a formal meta-analytic study.

Pedagogical Mechanisms of Guided Inquiry in Enhancing HOTS

The effectiveness of Guided Inquiry in developing Higher Order Thinking Skills (HOTS) can be explained by five interrelated and synergistic pedagogical mechanisms:

1. Cognitive Activation through Productive Questioning: The orientation phase of Guided Inquiry employs essential questions that generate cognitive conflict, prompting students to activate prior knowledge and identify knowledge gaps (Piaget, 1985). Questions such as "Why can heavier objects float?" or "How do plants make food without a mouth?" require inquiry rather than immediate answers. These questions create disequilibrium, motivating learners to engage in investigation toward cognitive equilibrium. They also function as cognitive scaffolds, focusing attention on relevant phenomena and activating appropriate mental schemas.
2. Tiered Scaffolding Aligned with the Student's ZPD: Guided Inquiry offers enough structure to prevent cognitive overload and productive failure, while also granting autonomy to promote self-regulated learning (Vygotsky, 1978). Teachers serve as facilitators who gradually fade scaffolding as students' competencies develop—from high support at the beginning of a topic to minimal

support at the end. Scaffolding may be procedural (e.g., steps of investigation), strategic (e.g., guiding questions), or conceptual (e.g., frameworks). Hmelo-Silver et al. (2007) emphasized that scaffolding does not diminish the cognitive benefits of inquiry; rather, it optimizes cognitive load and supports deeper learning.

3. Collaborative Knowledge Construction: During exploration and discussion phases, students negotiate meaning socially through argumentation, negotiation, and consensus building—processes that promote elaborative rehearsal and deep cognitive processing (Vygotsky, 1978). Explaining ideas to peers, defending claims with evidence, and reconciling conflicting viewpoints require engagement in higher-order cognitive processes. Collaborative inquiry also exposes students to multiple perspectives, enhancing cognitive flexibility. Research has shown that explaining to others is one of the most effective strategies for developing deep understanding.
4. Authentic Assessment and Metacognitive Reflection: Guided Inquiry integrates assessment for learning, not merely assessment of learning. Students regularly reflect on their thinking through journals, peer assessments, self-assessment rubrics, and reflective discussions. Prompts like "What worked in our investigation?" or "If we were to repeat this, what would we change?" help students identify effective strategies and regulate their learning—hallmarks of a self-directed learner (Kuhn & Dean, 2004). Such metacognitive awareness is transferable to other problem-solving contexts, making GI especially powerful for cultivating broadly applicable thinking skills.
5. Experiential Learning and Knowledge Transfer: Guided Inquiry offers concrete experiences that facilitate the transfer of concepts from classroom learning to real-world applications. Kolb's experiential learning cycle (1984)—comprising concrete experience, reflective observation, abstract conceptualization, and active experimentation—is naturally embedded in GI phases. Students learn not just *about* science, but *through* doing science, leading to deeper cognitive encoding and more accessible memory traces. Research on transfer suggests that knowledge acquired through active, meaningful experience is more transferable than passively acquired knowledge.

Sensitivity Analysis and Robustness of Findings

To assess the stability of the conclusions, sensitivity analyses were conducted by excluding studies with potential bias or lower methodological quality.

- Scenario 1: Excluding the study with marginal significance ($p = 0.066$, $n = 1$) did not affect the critical thinking effect size ($g = 1.33$), indicating minimal contribution from that study.
- Scenario 2: Excluding four studies lacking documented instrument validity yielded the following effect size ranges: critical thinking ($g = 1.28$–1.35), creative thinking ($g = 1.05$–1.12), and problem solving ($g = 1.27$–1.33), compared to the main analysis values of 1.33, 1.10, and 1.31 respectively.
- Scenario 3: Excluding studies with small sample sizes ($n < 30$, $n = 3$) produced similar results: critical thinking ($g = 1.30$–1.36), creative thinking ($g = 1.08$–1.13), and SPS N-Gain = 0.68–0.72 (vs. 0.70 in main analysis).

In all scenarios, the 95% confidence intervals did not include zero, indicating that the main conclusions are robust to the exclusion of potentially problematic studies. Although effect sizes slightly decreased by 2–5% on average, they remained in the "large to very large" category according to Cohen's benchmarks. This provides reasonable confidence that the overall findings were not driven by outliers or low-quality evidence.

Research Limitations and External Validity

Several limitations should be considered when interpreting the findings of this systematic review:

1. Unexplored Heterogeneity and Moderators: The reviewed studies exhibited considerable heterogeneity in multiple dimensions: (a) intervention duration ranged from 4 to 16 sessions (median = 8); (b) the degree of scaffolding varied from high structure to minimal guidance; (c) implementation settings spanned urban and rural schools, with differing access to laboratory resources; and (d) teacher profiles ranged from those trained in inquiry pedagogy to those using conventional instructional approaches. Quantitative moderator analysis could not be conducted due to insufficient disaggregated data in the primary studies. Key questions such as whether longer interventions (12+ sessions) yield better outcomes than shorter ones (4–6 sessions), or whether teacher training intensity moderates Guided Inquiry effectiveness, remain unanswered.
2. Publication Bias and the File Drawer Problem: Approximately 75% of the included studies were published in SINTA 2–5 journals in Indonesia, with a dominance of studies conducted in urban institutions and model schools. A potential publication bias is suggested by the fact that all primary studies reported positive effects of Guided Inquiry—none reported null or negative findings. Statistically, this is unlikely if Guided Inquiry were truly universally effective, indicating a possible file drawer problem, where studies with non-significant or counter-intuitive results are less likely to be submitted or published. A funnel plot from a prior meta-analysis (Cahaya et al., 2024) showed slight asymmetry in the small-study region, although Egger's test was not significant (p = 0.12). Trim-and-fill analysis suggested that if publication bias were accounted for, the overall effect size could decrease by 10–15%, though it would still remain within the "large effect" category.
3. Limitations in External Validity: Caution is needed when generalizing these findings to contexts beyond those studied. *(a) Geographic limitation*: The majority of studies were conducted in Java and Sumatra, with limited representation from Eastern Indonesia, where infrastructure, teacher preparation, and language barriers may differ significantly. *(b) School type limitation*: Most studies were conducted in urban schools with adequate infrastructure (labs, technology), which may not represent the roughly 60% of Indonesian schools in rural or semi-rural areas with limited resources. *(c) Student population limitation*: None of the studies involved students with special needs (e.g., learning disabilities, gifted students, non-native Indonesian speakers), leaving the applicability of Guided Inquiry to diverse learner populations uncertain. *(d) Subject matter limitation*: Most research focused on STEM subjects, with minimal evidence regarding the effectiveness of Guided Inquiry in social sciences or humanities.
4. Design and Measurement Limitations. *(a) Study design*: Only 25% of studies used true experimental designs with random assignment; the rest were quasi-

experimental, making them more vulnerable to selection bias and confounding. *(b) Posttest-only design*: About 60% of studies employed posttest-only designs, which fail to control for baseline differences, making it difficult to isolate treatment effects. *(c) Outcome instruments*: The wide range of HOTS assessment tools (a mix of standardized and self-developed instruments) hinders direct comparisons. Some instruments had only marginal reliability (α = 0.70–0.75), which may introduce measurement error. *(d) Instrument type*: Around 65% of studies used self-developed instruments. While validated, these may lack the rigor and generalizability of standardized tests and may be more sensitive to researcher expectations.

5. Lack of Long-Term Follow-Up and Transfer Evaluation: None of the studies evaluated the long-term retention of HOTS or the transfer of learning beyond six months post-intervention. All outcomes were measured either immediately or within two weeks after the intervention. Critical questions remain unanswered: Are HOTS gains from Guided Inquiry sustained over time? Do students transfer their inquiry skills to other domains or novel situations? Prior research on educational interventions indicates that immediate gains often diminish over time without continued support, thus the sustainability of Guided Inquiry's effects remains uncertain.
6. Hawthorne and Novelty Effects: Some of the high observed effects may be partially attributable to the Hawthorne effect (participants perform better when observed or aware they are being studied) or novelty effect (increased enthusiasm from trying something new). Without comparisons to long-term implementation or studies with delayed posttests, it is difficult to distinguish between genuine instructional effects and temporary performance boosts.

Alternative Explanations for the Findings

The high effectiveness of the Guided Inquiry model—especially in junior and senior high school contexts—should be interpreted with consideration of several alternative explanations:

1. Teacher Quality and Selection Bias: Studies showing the largest effects often involved teachers who were highly motivated, well-trained in inquiry pedagogy, and affiliated with model schools or universities. These teachers may have superior pedagogical content knowledge, better classroom management skills, and greater access to resources—factors independent of the instructional model itself. In other words, *"it's not just the model, it's the teacher implementing the model."* Future studies should include teacher quality indicators and involve multiple teachers per condition to disentangle model effects from teacher effects.
2. Resource and Infrastructure Advantages: Schools in the reviewed studies often had (a) adequate lab facilities; (b) manageable class sizes (average 28–34 students vs. 40+ in many schools); (c) access to technology (e.g., projectors, computers, internet); and (d) supportive administrative environments. These resources facilitate the successful implementation of Guided Inquiry but are not representative of most Indonesian schools. As such, the model's effectiveness in low-resource settings remains largely unknown.
3. Student Selection and Motivation: Some studies were conducted in accelerated or science-stream classes (MIPA/IPA), with students who were already high-

achieving and more intrinsically motivated toward science. Pre-existing differences in cognitive abilities, background knowledge, and motivation may have contributed to the large observed effects, particularly in quasi-experimental designs that lacked proper baseline control.

4. Measurement Sensitivity and Alignment: Outcome instruments were often closely aligned with the inquiry processes taught in the interventions (e.g., hypothesis formulation, data interpretation), creating a possible "teaching to the test" effect. When instruction and assessment are tightly aligned, improvements may reflect training in specific tasks rather than generalizable thinking abilities. Assessments using far-transfer measures—requiring students to apply inquiry skills in completely novel contexts—would provide stronger evidence of true HOTS development.
5. Comparison Condition Quality: Large effect sizes may also reflect weak comparison conditions. In many studies, the “conventional teaching” or “traditional lecture” used in control groups was poorly described or inconsistently implemented, making it an unfair comparator. Comparisons with other active learning strategies—such as problem-based learning, project-based learning, or collaborative problem-solving—would provide a more rigorous test of Guided Inquiry’s relative effectiveness.

Recognizing these alternative explanations does not diminish the value of the Guided Inquiry model. Instead, they provide a more nuanced understanding of the conditions under which the model is most effective, as well as the limitations of generalizing the findings across different contexts. Future research should aim to test these boundary conditions systematically and explore Guided Inquiry’s scalability and sustainability under realistic classroom constraints.

CONCLUSION

This systematic literature review of 20 studies—including 11 experimental or quasi-experimental designs, 4 meta-analyses, and 5 systematic reviews—provides compelling evidence for the effectiveness of the Guided Inquiry (GI) model in developing Higher Order Thinking Skills (HOTS) and Science Process Skills (SPS) in Indonesian science education. Meta-analyses reported very large and statistically significant effect sizes for critical thinking ($g = 1.33$), creative thinking ($g = 1.10$), and problem solving ($g = 1.31$). Experimental studies further confirmed substantial gains in SPS (N-Gain = 0.7) and science literacy (85–95% mastery), reinforcing the model's capacity to foster both cognitive and scientific competencies.

The effectiveness of GI was consistently high at the junior and senior high school levels (effect sizes between 0.8 and 1.3) when supported by structured scaffolding and appropriate durations (6–8 sessions). At the elementary level, outcomes varied more widely (effect sizes 0.4–1.10), depending on the intensity of scaffolding, the use of visual-concrete media, and implementation length. Notably, integration with technology and STEM contexts produced even greater improvements, with effect sizes increasing by approximately 0.15–0.25 points. These gains are underpinned by pedagogical mechanisms including cognitive activation, scaffolded learning within the Zone of Proximal Development (ZPD), collaborative knowledge construction, authentic assessment with reflective practice, and experiential learning that supports transfer.

However, the review also identified key limitations. These include high heterogeneity across studies, potential publication bias, limited external validity beyond urban and well-resourced schools, and a dominance of quasi-experimental designs without long-term follow-up. Additionally, the absence of a pre-registered review protocol and the underrepresentation of diverse learner populations constrain generalizability. As such, the findings are most applicable to junior and senior high school science education settings with trained teachers, adequate infrastructure, and sufficient instructional time. Future research should explore implementation in under-resourced contexts, include longitudinal evaluation, and assess effectiveness across broader subject areas and student populations.

RECOMMENDATION

To support effective implementation of the Guided Inquiry model in Indonesian science education, educators and school leaders are encouraged to adopt a phased and adaptive approach. Scaffolding should be tailored by educational level—intensive and explicit in elementary school, moderate and phased in junior high, and minimal in senior high as students transition toward open inquiry. High-quality implementation requires careful planning: teachers should receive professional development focused on facilitating inquiry (e.g., questioning techniques and managing productive struggle), allocate 6–8 sessions per topic, and design structured student worksheets (LKPD). Strategic integration with digital tools, STEM or socio-scientific contexts, and guided collaborative structures can further enhance engagement and cognitive outcomes. Continuous reflection and instructional adjustment—based on student learning evidence—should be embedded into teaching practice and shared within Professional Learning Communities (PLCs).

At the policy level, systemic support is essential. Education authorities should ensure curriculum flexibility, allowing time for in-depth inquiry learning through selective content coverage. Professional development must move beyond one-off workshops to sustained models, including intensive initial training, ongoing coaching, and PLC-based peer support. Investments should also be made in validated HOTS/SPS assessment banks, accessible repositories of Guided Inquiry-based LKPDs, and virtual labs or simulations. Assessment reforms are equally critical: standardized tests should incorporate inquiry and HOTS components, with complementary use of performance-based assessments and portfolios to capture deeper learning. To inform ongoing policy refinement, it is recommended that education ministries fund longitudinal research and experimental studies (RCTs or well-designed quasi-experiments) in real-world school settings.

For education researchers, addressing existing methodological gaps is a priority. Future studies should employ true experimental designs with random assignment, engage multiple teachers per condition, utilize validated instruments, and ensure adequate sample sizes based on power analysis. There is also a need to systematically examine critical moderators such as optimal scaffolding intensity, intervention duration thresholds, and effective models for technology integration through factorial designs. Research must expand to underrepresented contexts—rural, resource-constrained, and multilingual classrooms—while developing context-appropriate adaptations of the model. Scholars are also encouraged to shift focus from efficacy to implementation and scalability, identifying adoption barriers and sustainable delivery models. Transparent reporting, adherence to guidelines like CONSORT or STROBE, effect size reporting with confidence intervals, and the publication of null findings will strengthen the evidence base and guide more reliable educational decision-making.

ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to all researchers whose work contributed to this review, as well as to the institutions and databases that provided access to the publications analyzed.

REFERENCES

Adauyah, R., & Aznam, N. (2024). Guided inquiry learning model in chemistry education: A systematic review. *Jurnal Penelitian Pendidikan IPA (JPPIPA)*, *10*(3). https://doi.org/10.29303/jppipa.v10i3.6373

Ariskafitriani, I., & Tirtoni, F. (2025). Pengaruh efektivitas model pembelajaran inkuiri terbimbing melalui pendekatan kontekstual terhadap kemampuan berpikir kritis siswa. *JPDI (Jurnal Pendidikan Dasar Indonesia)*, *10*(2), 139–150. https://doi.org/10.26737/jpdi.v10i2.6942

Bybee, R. W. (1997). Achieving scientific literacy: From purposes to practices. *The Science Teacher*, *64*(3), 28–33.

Cahaya, I. M. E., Suryaningsih, N. M. A., Parwata, I. M. Y., & Poerwati, C. E. (2024). The influence of a guided inquiry learning model in improving students' creative thinking abilities: A meta-analysis study. *Indonesian Journal of Educational Development (IJED)*, *5*(3), 376–384. https://doi.org/10.59672/ijed.v5i3.4211

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dewanto, Santosa, T. A., Ratih, A., Asrizal, & Hardeli. (2024). The influence of the STEM-based guided inquiry model on students' creative thinking skills in science learning: A meta-analysis study. *Jurnal Penelitian Pendidikan IPA (JPPIPA)*, *10*(3), 88–95. https://doi.org/10.29303/jppipa.v10i3.6777

Doyan, A., Susilawati, & Hardiyansyah, A. (2021). Development of natural science learning tools with guided inquiry model assisted by real media to improve students' scientific creativity and science process skills. *Jurnal Penelitian Pendidikan IPA (JPPIPA)*, *7*(1), 15–20. https://doi.org/10.29303/jppipa.v7i1.485

Dykstra, D., Boyle, C., & Monarch, I. (1992). Studying conceptual change in learning physics. *Science Education*, *76*(6), 615–652. https://doi.org/10.1002/sce.3730760605

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The California Academic Press.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, *82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Hake, R. R. (1999). *Analyzing change/gain scores*. American Educational Research Association's Division D, Measurement and Research Methodology. https://web.physics.indiana.edu/sdi/AnalyzingChange-Gain.pdf

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking." *Journal of Educational Psychology*, *98*(2), 307–326. https://doi.org/10.1037/0022-0663.98.2.307

Hastuti, D. A. W., & Wiyanto. (2019). Pengaruh model pembelajaran guided inquiry dengan metode eksperimen terhadap keterampilan proses sains siswa. *Unnes Physics Education Journal (UPEJ)*, *8*(3), 289–298. http://journal.unnes.ac.id/sju/index.php/upej/article/view/30461

Hendra, Z., & Kurniati, N. (2024). Development of interactive learning multimedia based on guided inquiry to improve student learning outcomes in science content (force, motion, and energy transfer). *Fondatia: Jurnal Pendidikan Dasar*, *8*(1), 32–42. https://doi.org/10.36088/fondatia.v8i1.4477

Kim, H., & Im, S. (2021). Pre-service physics teachers' beliefs about learning physics and their learning achievement in physics. *Asia-Pacific Science Education*, *7*(2), 500–521. https://doi.org/10.1163/23641177-bja10038

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Kuhn, D., & Dean, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into Practice*, *43*(4), 268–273. https://doi.org/10.1207/s15430421tip4304_4

Kusumaningsih, S. F., & Trimulyono, G. (2020). Pengembangan LKPD berbasis guided inquiry untuk melatihkan keterampilan literasi sains pada materi bakteri kelas X SMA. *BioEdu: Berkala Ilmiah Pendidikan Biologi*, *9*(3), 378–380. https://ejournal.unesa.ac.id/index.php/bioedu

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Maloney, D. P., O'Kuma, T., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(S1), S12–S23. https://doi.org/10.1119/1.1371296

Maulidiyah, N., & Wulandari, F. (2023). The effect of the guided inquiry model on the reasoning abilities of elementary school students. *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme*, *5*(2), 653–668. https://doi.org/10.37680/scaffolding.v5i2.3247

Nur, J., Utami, S., Nuraini, H., & Kurniawan, A. (2024). Meta-analysis influence of integrated mind mapping inquiry based learning model on student problem solving skills. *Jurnal Penelitian Pendidikan IPA (JPPIPA)*, *10*(3), 1361–1371. https://doi.org/10.29303/jppipa.v10i3.6804

Nurkhasanah, A. F., Sabri, A., & Nelwati, S. (2024). Penerapan model inkuiri terbimbing sebagai solusi keaktifan peserta didik di SMP N 01 Tiumang Dharmasraya. *Islamika: Jurnal Keislaman dan Ilmu Pendidikan*, *6*(3), 1474–1483.

OECD. (2019). *PISA 2018 results (Volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463–466. https://doi.org/10.1126/science.1183944

Pahriah, Adnyana, P. B., Ariawan, I. P. W., & Wesnawa, I. G. A. (2024). Effectiveness of the constructivist approach (guided inquiry) in chemistry learning: A systematic review. *Hydrogen: Jurnal Kependidikan Kimia*, *12*(5), 1152–1165. https://doi.org/10.33394/hjkk.v12i5.13353

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. University of Chicago Press.

Pikoli, M. (2020). Using guided inquiry learning with multiple representations to reduce misconceptions of chemistry teacher candidates on acid-base concept. *International Journal of Active Learning*, *5*(1), 1–10. http://journal.unnes.ac.id/nju/index.php/ijal

Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton University Press.

Qoyyimah, T. F., & Nugroho, O. F. (2021). Pengaruh model pembelajaran guided inquiry berbasis pictorial riddle dalam meningkatkan berpikir kreatif siswa pada pembelajaran IPA di SDN Gudang. *Jurnal PERSEDA*, *4*(3), 141–147. https://jurnal.ummi.ac.id/index.php/perseda

Sakina, B. S., Ekawati, R., & Dasna, I. W. (2024). The influence of guided inquiry worksheets on science literacy and learning outcomes in fourth-grade students. *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme*, *6*(3), 123–136. https://doi.org/10.37680/scaffolding.v6i3.6376

Samadun, Setiani, R., Dwikoranto, & Marsini. (2023). Effectiveness of inquiry learning models to improve students' critical thinking ability. *International Journal of Recent Educational Research*, *4*(2), 203–212. https://journal.ia-education.com/index.php/ijorer

Sinta, A. D., & Agustina, P. (2024). Science process skills and biology learning outcomes of high school students through the application of the guided inquiry learning model. *Edubiotik: Jurnal Pendidikan, Biologi dan Terapan*, *9*(01), 45–53. https://doi.org/10.33503/ebio.v9i01.4021

Suryono, W., Winiasri, L., Santosa, T. A., Sappaile, B. I., & Solehuddin, M. (2023). Effectiveness of the inquiry training model to improve students' critical thinking skills in learning: Systematic literature reviews and meta-analysis. *Jurnal Penelitian Pendidikan IPA (JPPIPA)*, *9*(10), 742–750. https://doi.org/10.29303/jppipa.v9i10.4804

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Yulianis, F., & Mawardi, M. (2021). Pengaruh penggunaan sistem pembelajaran flipped-guided inquiry learning (FGIL) terhadap hasil belajar siswa pada materi laju reaksi. *Entalpi Pendidikan Kimia*, *2*(2), 1–8. https://doi.org/10.1021/bk-2019-1336.ch013

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, *12*(2), 145–181. https://doi.org/10.1207/S15327809JLS1202_1