

Pengaruh Chatbot Scaffolding Sokratik Berbasis AI terhadap Penalaran Ilmiah Siswa Sekolah Dasar: Kuasi-Eksperimental Model Inkuiri 5E

Nafi Isbadrianingtyas^{*1}, Alfarid Hendro Yuwono², Usman Nurhasan³, Norazah Mohd Nordin⁴

¹Departemen Pendidikan Guru Sekolah Dasar, Fakultas Ilmu Pendidikan, Universitas Negeri Surabaya, Indonesia

²Departemen Teknik Listrik, Fakultas Vokasi, Universitas Negeri Surabaya, Indonesia

³Departemen Teknologi Informasi, Politeknik Negeri Malang, Indonesia

⁴Department of Higher Education, Ministry of Higher Education, Universiti Kebangsaan Malaysia

Email: ¹nafiisbadrianingtyas@unesa.ac.id, ²alfaridyuwono@unesa.ac.id,

³usmannurhasan@polinema.ac.id, ⁴drnmn@ukm.edu.my

Abstrak

Penalaran ilmiah siswa sekolah dasar di Indonesia masih rendah; skor sains PISA 2022 hanya 383 (peringkat 67 dari 81 negara) dengan sekitar 34% siswa mencapai kompetensi minimum, sementara keterbatasan waktu guru menyulitkan pemberian scaffolding dialogis yang adaptif secara individual di kelas besar. Penelitian ini bertujuan menguji efektivitas chatbot Sokratik berbasis AI sebagai instrumen scaffolding dialogis dalam model inkuiri 5E untuk meningkatkan penalaran ilmiah siswa kelas 5 sekolah dasar. Menggunakan desain kuasi-eksperimental pretest-posttest non-equivalent control group pada 60 siswa (30 eksperimen, 30 kontrol) dari dua kelas paralel, intervensi chatbot berbasis aturan (rule-based) diberikan selama enam minggu (12 sesi) pada fase Explore dan Explain; data dianalisis menggunakan ANCOVA dan pengkodean Toulmin's Argument Pattern (TAP). Hasil menunjukkan kelompok eksperimen mengalami peningkatan penalaran ilmiah yang signifikan secara statistik ($F(1,57) = 28,45$; $p < 0,001$) dengan ukuran efek besar ($\eta^2 = 0,333$); 74% episode dialog menghasilkan penjelasan berbasis bukti secara mandiri, dan miskonsepsi persisten berkurang 46%. Temuan ini memberikan bukti empiris bahwa chatbot Sokratik berbasis AI dapat berperan sebagai mediator epistemis yang memperkuat koordinasi teori-bukti sekaligus menawarkan model scaffolding individual yang skalabel bagi kelas besar dengan keterbatasan sumber daya guru.

Kata kunci: chatbot berbasis AI, model inkuiri 5E, penalaran ilmiah, pendidikan dasar, scaffolding Sokratik.

Effect of an AI-Based Socratic Scaffolding Chatbot on Elementary Students' Scientific Reasoning: A Quasi-Experimental Study within the 5E Inquiry Model

Abstract

Elementary students' scientific reasoning in Indonesia remains low; the PISA 2022 science score was only 383 (67th of 81 countries), with about a third of students reaching minimum proficiency, while teachers' limited time constrains adaptive, individualized dialogic scaffolding in large classrooms. This study aims to examine the effectiveness of an AI-based Socratic chatbot as a dialogic scaffolding tool within the 5E inquiry model for enhancing the scientific reasoning of fifth-grade students. A quasi-experimental pretest-posttest non-equivalent control group design was employed with 60 students (30 experimental, 30 control) from two parallel classes; a rule-based chatbot intervention was delivered over six weeks (12 sessions) during the Explore and Explain phases, with data analyzed using ANCOVA and Toulmin's Argument Pattern (TAP) coding. Results show the experimental group achieved significantly higher post-test scores ($F(1,57) = 28.45$; $p < 0.001$) with a large effect size ($\eta^2 = 0.333$); 74% of dialogue episodes produced independently constructed evidence-linked explanations, and persistent misconceptions decreased by 46%. These findings provide empirical evidence that AI-based Socratic chatbots can function as epistemic mediators that strengthen theory-evidence coordination while offering a scalable model of individualized scaffolding for resource-constrained, large classrooms.

Keywords: AI-powered chatbot, 5E inquiry model, elementary education, scientific reasoning, Socratic scaffolding.

1. PENDAHULUAN

Kemampuan penalaran ilmiah merupakan salah satu indikator kunci literasi sains yang menjadi perhatian global. Namun demikian, capaian literasi sains siswa Indonesia usia 15 tahun pada Programme for International Student Assessment (PISA) 2022 hanya mencapai skor rata-rata 383, menempatkan Indonesia pada peringkat ke-67 dari 81 negara peserta, dengan hanya sekitar 34% siswa yang mencapai level kompetensi minimum (Level 2), jauh di bawah rata-rata negara-negara OECD [1]. Kondisi ini menunjukkan bahwa persoalan penalaran ilmiah bukan sekadar isu teoretis, melainkan tantangan empiris mendesak dalam sistem pendidikan dasar di Indonesia yang memerlukan intervensi pedagogis konkret sejak jenjang sekolah dasar.

Kemampuan bernalar secara ilmiah merupakan pilar fundamental dalam literasi sains dan keberhasilan akademik siswa di masa depan. Penalaran ilmiah bukan sekadar fenomena perkembangan kognitif alami, melainkan sebuah praktik epistemik konstruktif di mana siswa belajar membangun hipotesis, menyelaraskan bukti dengan teori, dan merumuskan penjelasan kausal atas fenomena alam[2][3]. Pada tingkat sekolah dasar, perkembangan ini terjadi secara bertahap melalui lintasan belajar (*learning trajectories*), di mana siswa kelas atas (4–6) mulai bertransisi dari pemikiran deskriptif menuju koordinasi teori-bukti yang lebih kompleks[4][5].

Namun, perkembangan kognitif ini tidak terjadi secara otomatis atau universal. Banyak siswa menghadapi hambatan konseptual signifikan yang membatasi kemampuan mereka untuk mencapai level penalaran yang lebih tinggi[6]. Di sekolah dasar, miskonsepsi sering kali muncul dalam berbagai domain; misalnya, siswa sering menganggap gaya sebagai sifat internal benda atau menyamakan panas dengan suhu[7][8]. Hambatan serupa ditemukan pada domain biologi dan geografi, terutama saat siswa diminta memahami variabel yang tidak kasat mata seperti arus energi atau sistem Bumi yang terintegrasi [9][10]. Tanpa intervensi yang tepat, siswa cenderung terjebak pada pemikiran permukaan dan gagal memahami ide-ide besar sains (*Big Ideas of Science*)[11].

Secara teoretis, perubahan kognitif yang positif memerlukan bimbingan (*scaffolding*) yang adaptif, kontinu, dan selaras dengan *Zone of Proximal Development* (ZPD) siswa[12][13]. Sayangnya, di banyak ruang kelas sekolah dasar, guru sering kali kesulitan memberikan bimbingan individual yang fokus dan tepat waktu karena keterbatasan waktu dan sumber daya[14]. Meskipun kemajuan Kecerdasan Buatan (AI) menawarkan peluang besar untuk menyediakan *scaffolding* dialogis yang adaptif, penelitian empiris mengenai penggunaan AI untuk mendukung koordinasi teori-bukti pada siswa sekolah dasar masih sangat terbatas[15][16].

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengintegrasikan chatbot Sokratik berbasis AI ke dalam model pembelajaran inkuiri 5E. Berbeda dengan alat pemberi informasi konvensional, chatbot ini dirancang sebagai mediator dialogis yang memicu ketegangan kognitif produktif melalui pertanyaan reflektif. Melalui pendekatan ini, penelitian ini berupaya menjawab apakah interaksi dialogis dengan AI dapat secara signifikan meningkatkan penalaran ilmiah siswa dan membantu mereka merekonstruksi miskonsepsi menjadi pemahaman yang selaras secara saintifik.

Kebaruan (*novelty*) penelitian ini terletak pada tiga aspek. Pertama, penelitian ini secara eksplisit mengaitkan tiga jenis bukti — hasil psikometrik model Rasch, hasil inferensial ANCOVA, dan analisis proses dialog berbasis Toulmin's Argument Pattern (TAP) — dalam satu kerangka evaluasi dampak AI terhadap penalaran ilmiah, yang jarang dilakukan secara terintegrasi pada penelitian *scaffolding* AI di jenjang sekolah dasar. Kedua, penelitian ini memosisikan chatbot Sokratik sebagai agen berbasis aturan (*rule-based*) yang secara spesifik dirancang untuk fase *Explore* dan *Explain* pada model 5E, berbeda dari sebagian besar studi terdahulu yang menerapkan chatbot generatif secara umum tanpa penahapan inkuiri eksplisit. Ketiga, penelitian ini memberikan bukti kontekstual dari kelas sekolah dasar di Indonesia, mengisi kesenjangan data empiris pada populasi usia dini yang selama ini didominasi oleh studi pada jenjang pendidikan menengah dan tinggi. Dengan demikian, studi ini memberikan kontribusi pada pengembangan pedagogi AI yang responsif dalam ekosistem pendidikan sains masa kini.

2. METODE PENELITIAN

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain kuasi-eksperimental (*quasi-experimental design*), khususnya model *Pre-test and Post-test Non-equivalent Control Group Design* [17]. Desain ini dipilih untuk mengevaluasi dampak intervensi chatbot Sokratik terhadap penalaran ilmiah siswa dalam kondisi kelas yang sudah terbentuk secara alami tanpa melakukan randomisasi individu.

2.2. Partisipan, Konteks, dan Teknik Sampling

Subjek penelitian terdiri atas 60 siswa kelas 5 sekolah dasar yang berasal dari dua kelas paralel. Pemilihan kelas dilakukan dengan teknik *convenience sampling* atas dua kelas paralel yang telah terbentuk secara alami (*intact group*) pada sekolah tempat penelitian dilaksanakan, sedangkan penentuan status eksperimen dan kontrol

pada kedua kelas tersebut ditentukan melalui pengundian acak sederhana (simple random assignment) di tingkat kelas untuk meminimalkan bias seleksi sistematis, mengingat randomisasi pada tingkat individu tidak dimungkinkan karena struktur kelas yang sudah terbentuk sebelumnya. Satu kelas ditetapkan sebagai kelompok eksperimen ($n=30$), yang mengintegrasikan chatbot Sokratik berbasis AI dalam pembelajarannya, sementara kelas lainnya bertindak sebagai kelompok kontrol ($n=30$) yang mengikuti pembelajaran inkuiri konvensional tanpa bantuan AI.

2.3. Prosedur Intervensi: Integrasi AI dalam Model 5E

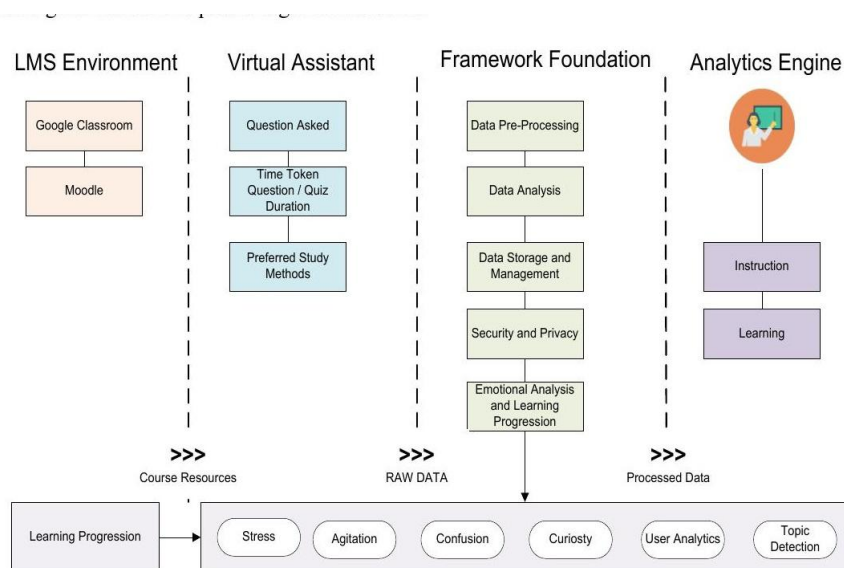
Intervensi dilaksanakan selama enam minggu dengan frekuensi dua sesi per minggu (total 12 sesi), masing-masing berdurasi 35–40 menit, mengikuti jadwal pembelajaran IPA reguler di sekolah. Kedua kelompok mengikuti fase pembelajaran berbasis model 5E (Engage, Explore, Explain, Elaborate, Evaluate). Namun, pada kelompok eksperimen, chatbot Sokratik diintegrasikan secara spesifik pada fase Explore dan Explain:

- Fase Explore: Siswa berinteraksi dengan chatbot untuk mengeksplorasi data hasil observasi. AI memberikan pertanyaan pemantik yang mendorong siswa mengaitkan bukti dengan fenomena.
- Fase Explain: Chatbot memberikan bimbingan dialogis (Sokratik) untuk membantu siswa merumuskan penjelasan ilmiah tanpa memberikan jawaban secara langsung, guna memicu kemandirian berpikir dalam ZPD mereka.

2.4. Arsitektur dan Model AI Chatbot Sokratik

Chatbot Sokratik yang digunakan dalam penelitian ini merupakan sistem berbasis aturan (rule-based dialogue system), bukan model bahasa besar generatif (generative large language model/LLM) seperti GPT. Sistem dibangun di atas mesin klasifikasi intent berbasis kata kunci (keyword-based intent classification) yang memetakan respons siswa ke dalam kategori Klaim, Data, atau Penjamin berdasarkan skema Toulmin's Argument Pattern, kemudian memicu bank pertanyaan Sokratik berjenjang (pertanyaan klarifikasi, pertanyaan sangkalan/refutational prompting, dan pertanyaan elaborasi) yang telah disusun dan divalidasi oleh pakar pendidikan IPA sesuai tahapan fase Explore dan Explain pada model 5E. Mekanisme adaptif diterapkan melalui pohon keputusan (decision tree) yang menyesuaikan tingkat kesulitan pertanyaan berikutnya berdasarkan kualitas argumen siswa pada giliran (turn) sebelumnya, tanpa memberikan jawaban secara langsung. Pendekatan rule-based dipilih untuk menjaga prediktabilitas dan keamanan konten bagi pengguna anak usia sekolah dasar, sekaligus menghindari risiko halusinasi jawaban yang umum ditemukan pada model generatif. Sebagaimana ditunjukkan pada Gambar 1, chatbot terintegrasi dengan platform Learning Management System (LMS) yang lazim digunakan sekolah, yaitu Google Classroom dan Moodle, dengan modul Analytics Engine yang merekam durasi interaksi dan pola pertanyaan siswa untuk keperluan analisis proses.

2.5. Instrumen Penelitian



Gambar 1. Arsitektur konseptual ekosistem pembelajaran yang didukung AI

Dua instrumen utama digunakan dalam penelitian ini:

- Tes Penalaran Ilmiah: Terdiri dari soal-soal pilihan ganda beralasan yang telah divalidasi dan diuji reliabilitasnya menggunakan model Rasch[18][19]. Soal ini mengukur kemampuan siswa dalam membuat klaim, mengidentifikasi bukti, dan memberikan pembenaran (reasoning)[20].

Sebelum uji coba lapangan, instrumen Scientific Reasoning Test (SRT) melalui tahap validasi isi (content validity) oleh panel tiga pakar (dua pakar pendidikan IPA sekolah dasar dan satu pakar psikometri) menggunakan indeks Aiken's V, dengan seluruh butir memperoleh $V \geq 0,78$ (kategori valid). Instrumen kemudian diujicobakan (pilot test) pada 20 siswa kelas 5 di luar sampel penelitian sebelum dilakukan kalibrasi akhir menggunakan model Rasch.

- Log Percakapan (Dialogue Logs): Data percakapan antara siswa dan chatbot direkam untuk dianalisis secara kualitatif guna memahami proses argumentasi siswa.

2.6. Teknik Analisis Data

Data kuantitatif dianalisis menggunakan statistik inferensial:

- ANCOVA (Analysis of Covariance): Digunakan untuk menguji perbedaan signifikan pada skor post-test antara kedua kelompok dengan mengontrol nilai pre-test sebagai kovariat [21].
- Effect Size (Partial Eta Squared - η^2): Dihitung untuk menentukan besarnya pengaruh intervensi, dengan interpretasi mengacu pada pedoman [22].
- Analisis Kualitatif: Log percakapan dianalisis menggunakan skema pengkodean Toulmin's Argument Pattern (TAP) untuk mengidentifikasi perkembangan struktur argumen siswa (Klaim, Bukti, Penjamin)[23].

Tabel 1. Konstruk, Variabel, Instrumen, dan Teknik Analisis yang Digunakan dalam Penelitian

Konstruk/Variabel	Peran Analitis	Sumber Data/Instrumen	Teknik Analisis
Penalaran Ilmiah (Konstruk Laten)	Variabel Terikat	Tes Penalaran Ilmiah (SRT), 15 butir soal bergambar diadaptasi dari CTSR	Pemodelan Rasch; ANCOVA; η^2
Strategi Kontrol Variabel (CVS)	Indikator Sub- konstruk	Sub-himpunan butir soal SRT	Inspeksi sub-skala Rasch
Penalaran Kausal- Proporsional	Indikator Sub- konstruk	Sub-himpunan butir soal SRT	Inspeksi sub-skala Rasch
Koordinasi Teori-Bukti	Indikator Sub- konstruk	Sub-himpunan butir soal SRT	Inspeksi sub-skala Rasch
Kondisi Instruksional	Variabel Bebas	Penugasan Eksperimen (AI-5E vs. Inkuiri Konvensional)	Faktor Kelompok dalam ANCOVA
Interaksi Scaffolding Sokratik	Variabel Proses	Log percakapan Siswa-AI	Analisis sekuensial TAP

3. HASIL DAN PEMBAHASAN

3.1. Analisis Awal dan Kualitas Pengukuran

Sebelum pengujian hipotesis, kualitas instrumen Scientific Reasoning Test (SRT) dievaluasi menggunakan model Rasch melalui perangkat lunak Winsteps[24]. Hasil diagnosis menunjukkan bahwa instrumen memiliki reliabilitas item yang sangat tinggi (0,92) dan reliabilitas person yang baik (0,84). Nilai rata-rata Infit dan Outfit MNSQ berada pada rentang 0,98 hingga 1,02, yang berarti seluruh item berfungsi secara konsisten tanpa gangguan (noise) yang signifikan. Verifikasi struktur internal instrumen melalui Principal Component Analysis (PCA) terhadap residu [25] menunjukkan bahwa dimensi laten utama penalaran ilmiah menjelaskan 46,8% varians data, sementara varians tak terjelaskan pada kontras pertama hanya sebesar 6,2% (di bawah ambang batas kritis 15%), yang mendukung syarat unidimensionalitas untuk analisis statistik parametrik. Uji Differential Item Functioning (DIF) juga tidak menemukan bias item yang signifikan antar kelompok [26].

Tabel 2. Ringkasan Statistik Model Rasch untuk Instrumen SRT (N=60)

Parameter Psikometrik	Nilai	Kriteria	Status
Rata-rata Infit MNSQ (Butir)	1,02	0,6–1,4	Diterima
Rata-rata Outfit MNSQ (Butir)	0,98	0,6–1,4	Diterima
Reliabilitas Butir Soal	0,92	> 0,80	Sangat Tinggi
Indeks Pemisahan Butir Soal	3,41	> 3,0	Sangat Baik
Reliabilitas Siswa	0,84	> 0,70	Baik
Indeks Pemisahan Siswa	2,28	> 2,0	Baik
Varians Dijelaskan oleh Ukuran (PCA)	46,8%	> 40%	Cukup
Varians Tak Terjelaskan, Kontras 1	6,2%	< 15%	Cukup

3.2. Karakteristik dan Pemeriksaan Baseline

Analisis awal dilakukan untuk memetakan kondisi dasar kedua kelompok sebelum intervensi [27]. Hasil uji t-independen menunjukkan perbedaan rata-rata skor pre-test yang sangat kecil (0,05 logit) dengan ukuran efek yang diabaikan ($d = 0,07$), membuktikan bahwa kelompok eksperimen dan kontrol memiliki kemampuan awal yang setara (Tabel 3). Selain itu, seluruh asumsi statistik untuk ANCOVA telah terpenuhi, termasuk normalitas residu ($p = 0,342$), homogenitas varians ($p = 0,289$), dan homogenitas kemiringan regresi ($p = 0,363$) [28].

Tabel 3. Perbandingan Skor Pre-test Antar Kelompok (N=60)

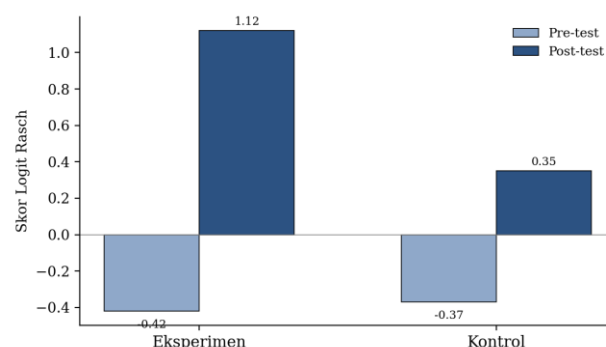
Kelompok	N	Rata-rata (Logit)	SD	p-value	Cohen's d
Eksperimen	30	-0,42	0,65	0,776	0,07
Kontrol	30	-0,37	0,71	–	–

3.3. Efek Chatbot Sokratik AI terhadap Penalaran Ilmiah

Analisis ANCOVA menunjukkan bahwa kondisi instruksional memiliki dampak signifikan terhadap keterampilan penalaran ilmiah siswa ($F(1,57) = 28,45$; $p < 0,001$) [29]. Kelompok eksperimen menunjukkan peningkatan skor sebesar +1,54 logit, sementara kelompok kontrol meningkat sebesar +0,72 logit, sebagaimana divisualisasikan pada Gambar 2. Ukuran efek parsial eta squared (η^2) sebesar 0,333 tergolong besar berdasarkan pedoman interpretasi umum untuk ANCOVA ($\eta^2 \geq 0,14$ dikategorikan besar) [22], yang berarti kondisi instruksional (chatbot Sokratik AI vs. inkuiri konvensional) menjelaskan sekitar 33,3% varians skor penalaran ilmiah siswa setelah mengontrol skor pre-test. Nilai ini melampaui ukuran efek rata-rata intervensi teknologi pendidikan pada umumnya yang cenderung berada pada kategori kecil hingga sedang, sehingga mengindikasikan dampak substansial dari scaffolding dialogis berbasis AI dibandingkan pendekatan inkuiri konvensional.

Tabel 4. Statistik Deskriptif Skor Logit Rasch Pre-test dan Post-test (N=60)

Kelompok	N	Rata-rata Pre-test (SD)	Rata-rata Post-test (SD)	Peningkatan Logit	Cohen's d
Eksperimen	30	-0,42 (0,65)	1,12 (0,58)	+1,54	Eksperimen
Kontrol	30	-0,37 (0,71)	0,35 (0,62)	+0,72	Kontrol



Gambar 2. Perbandingan Skor Logit Rasch Pre-test dan Post-test Antar Kelompok

3.4. Pola Sub-konstruk dan Interaksi Dialog Sokratik

Secara deskriptif, perbedaan skor terbesar antar kelompok ditemukan pada sub-konstruk Koordinasi Teori-Bukti (+1,03 logit) [30], selaras dengan fokus utama scaffolding Sokratik AI yang menggunakan pertanyaan interogatif untuk mendorong integrasi klaim dengan bukti. Sebaliknya, peningkatan pada penalaran kausal-proporsional lebih kecil (+0,57 logit), menunjukkan bahwa area ini memerlukan durasi intervensi yang lebih panjang.

Tabel 5. Distribusi Rata-rata Skor per Sub-konstruk (dalam Logit)

Sub-konstruk	Eksperimen (M)	Kontrol (M)	Perbedaan (Logit)
Strategi Kontrol Variabel (CVS)	1,25	0,45	+0,80
Penalaran Kausal-Proporsional	0,85	0,28	+0,57
Koordinasi Teori-Bukti	1,35	0,32	+1,03

Seluruh siswa dalam kelompok eksperimen memenuhi kriteria paparan dialogis, dengan rata-rata 18,4 putaran (turns) per sesi, rata-rata 12,6 kata per putaran siswa, dan 4,2 episode dialog per sesi [31]. Analisis log percakapan mengonfirmasi kepatuhan sistem terhadap metode Sokratik, di mana AI tidak memberikan jawaban langsung melainkan memicu proses berpikir [32].

3.5. Struktur Argumentasi dan Bukti Dialog

Analisis kualitatif mengidentifikasi bahwa pola penjelasan berbasis bukti (evidence-linked explanation) muncul pada 74% dari total 126 episode dialog yang dianalisis [33]. Struktur argumen dalam dialog didominasi oleh Klaim (42%; $f=284$), diikuti Data (32%; $f=216$), dan Penjamin/Warrant (26%; $f=172$) yang muncul setelah dilakukan pelacakan (probing) oleh sistem. Siswa terbukti mampu menghubungkan observasi empiris dengan klaim ilmiah secara kohesif melalui bimbingan AI, dan pertanyaan pelacak dari AI secara efektif memicu upaya siswa memberikan penjelasan mekanistik [34].

Data log dialog memberikan bukti nyata tentang bagaimana siswa merekonstruksi penjelasan mereka setelah menerima rangsangan refutasi (refutational prompting) dari AI [35][36]. Salah satu contoh menunjukkan siswa yang awalnya menganggap musim disebabkan oleh jarak Bumi ke Matahari, berhasil mengenali anomali tersebut setelah diberikan pertanyaan kritis oleh AI tentang perbedaan musim di belahan Bumi Utara dan Selatan, lalu membangun penjelasan baru berdasarkan kemiringan poros Bumi:

Siswa: "Musim panas adalah hasil dari Bumi yang bergerak lebih dekat ke Matahari." (Kesalahpahaman Awal)

AI: "Kalau begitu, mengapa belahan bumi selatan mengalami musim dingin sementara belahan bumi utara mengalami musim panas?" (Pertanyaan Sangkalan)

Siswa: "Oh, kalau begitu jaraknya tidak sama untuk semua bagian Bumi?" (Pengenalan Anomali)

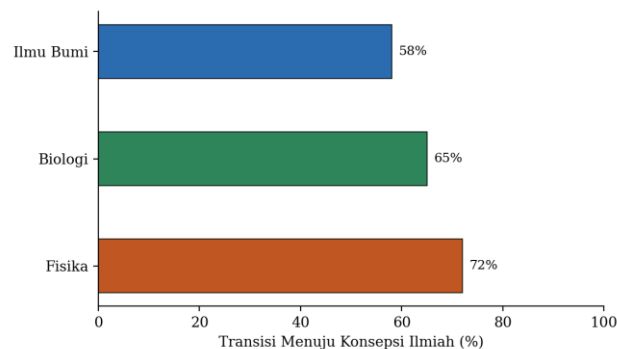
AI: "Perhatikan posisi kemiringan Bumi saat berputar mengelilingi Matahari dalam simulasi. Apa yang Anda amati?"

Siswa: "Bagian yang miring ke arah matahari menerima lebih banyak cahaya langsung... Jadi musim terjadi karena kemiringan tersebut." (Upaya Penjelasan yang Direkonstruksi)

Cuplikan di atas menggambarkan koeksistensi stimulus kontradiktif yang diberikan AI dan munculnya upaya penjelasan yang lebih rumit oleh siswa, yang secara konsisten selaras dengan perubahan skor pasca-tes SRT.

3.6. Pergeseran Konsepsi Sains Siswa Lintas Domain

In Intervensi ini berhasil mereduksi "Miskonsepsi Persisten" dari 64% menjadi 18% pada akhir penelitian (penurunan sebesar 46%), sementara kategori "Pemahaman Selaras secara Saintifik" meningkat sebesar 40%. Berdasarkan domain sains, sebagaimana ditunjukkan pada Gambar 3, transisi konseptual tertinggi terjadi pada bidang Fisika (72%; miskonsepsi "gaya sebagai sifat internal benda"), disusul Biologi (65%; "tumbuhan memperoleh makanan langsung dari tanah"), dan Ilmu Bumi (58%; "perubahan musim disebabkan jarak Bumi ke Matahari").



Gambar 3. Transisi Menuju Konsepsi Ilmiah Berdasarkan Domain Sains

3.7. Pembahasan: AI sebagai Mediator Inkuiri

Studi ini bergerak dari sekadar evaluasi efektivitas menuju pemeriksaan mendalam terhadap kondisi di mana AI dialogis memfasilitasi penalaran ilmiah dan perubahan konseptual dalam pembelajaran inkuiri [37]. Triangulasi hasil model Rasch, ANCOVA, dan analisis log percakapan menunjukkan bahwa peningkatan penalaran ilmiah siswa bukan terjadi secara terisolasi, melainkan hasil dari praktik dialog berbasis bukti yang berkelanjutan. Chatbot Sokratik berfungsi sebagai scaffold mirip More Knowledgeable Other (MKO) yang menjaga ketegangan kognitif produktif tetap berada dalam ZPD siswa, bertindak sebagai mediator interaksional melalui dukungan dialogis yang terdiferensiasi — sesuatu yang sulit dicapai dalam konteks kelas besar dengan bimbingan guru tunggal.

Ditinjau dari Cognitive Load Theory, chatbot Sokratik membantu mengurangi beban kognitif asing (extraneous load) melalui segmentasi penalaran. Dengan memecah urutan argumen Klaim–Data–Penjamin melalui dialog, siswa dapat berkonsentrasi pada satu bagian argumen dalam satu waktu, meningkatkan koherensi skema berpikir sekaligus mengurangi hambatan yang muncul akibat kesenjangan antara bukti dan penjelasan. Salah satu kontribusi penting penggunaan AI ini adalah pergeseran praktik epistemik siswa, dari sekadar mencari jawaban menjadi inkuiri yang digerakkan pertanyaan. Melalui strategi epistemic withholding dan refutational prompting, chatbot mendorong siswa mengevaluasi klaim berdasarkan bobot bukti, bukan otoritas, yang menunjukkan potensi AI dalam memupuk skeptisisme ilmiah dan penalaran kritis sejak pendidikan dasar.

3.8. Pembahasan: Implikasi Praktis

Secara praktis, temuan ini memiliki implikasi langsung bagi implementasi di sekolah. Pertama, integrasi chatbot Sokratik ke dalam Learning Management System (LMS) yang sudah lazim digunakan sekolah, seperti Google Classroom atau Moodle sebagaimana diilustrasikan pada arsitektur Gambar 1, memungkinkan adopsi tanpa memerlukan infrastruktur tambahan yang signifikan. Kedua, guru memerlukan pelatihan singkat (coaching clinic) mengenai cara memantau log percakapan siswa dan menindaklanjuti pola miskonsepsi yang teridentifikasi oleh Analytics Engine, agar peran chatbot tetap berfungsi sebagai pendamping — bukan pengganti — peran pedagogis guru. Ketiga, sekolah dengan rasio guru-siswa yang tinggi dapat memanfaatkan chatbot ini secara khusus pada fase Explore dan Explain untuk memberikan bimbingan individual yang selama ini sulit dijangkau oleh guru tunggal di kelas besar.

3.9. Pembahasan: Keterbatasan Penggunaan AI

Meskipun demikian, penerapan AI dalam pembelajaran anak usia sekolah dasar tetap perlu memperhatikan sejumlah keterbatasan. Pemilihan arsitektur rule-based pada penelitian ini secara inheren membatasi risiko halusinasi jawaban (hallucination) dan bias konten yang umum ditemukan pada model generatif, namun konsekuensinya adalah keterbatasan fleksibilitas dalam merespons pertanyaan siswa yang berada di luar bank pertanyaan yang telah dirancang. Apabila pengembangan lanjutan mengarah pada integrasi model bahasa generatif untuk memperluas fleksibilitas dialog, potensi risiko bias algoritmik dan halusinasi informasi ilmiah perlu dimitigasi melalui mekanisme pengawasan manusia (human-in-the-loop) dan pembatasan domain topik yang ketat. Selain itu, durasi intervensi yang terbatas pada penelitian ini turut memunculkan risiko ketergantungan (dependency) jangka panjang terhadap bimbingan AI apabila tidak diimbangi dengan penguatan kemandirian bernalar siswa secara bertahap; oleh karena itu, penelitian lanjutan disarankan mengukur secara eksplisit trajektori kemandirian siswa setelah scaffolding AI dihentikan (fading effect).

3.10. Keterbatasan Penelitian dan Implikasi Masa Depan

Temuan ini harus diinterpretasikan dalam batasan desain penelitian. Ukuran sampel yang terbatas dan konteks sekolah tunggal membatasi generalisasi temuan secara luas. Selain itu, keterkaitan erat antara chatbot dan desain instruksional 5E membuat atribusi kausal yang murni dari AI saja menjadi sulit dipisahkan. Namun, interdependensi ini justru mencerminkan kondisi autentik implementasi teknologi pendidikan, di mana agen AI tidak beroperasi dalam ruang pedagogis yang terisolasi. Penelitian selanjutnya disarankan menggunakan desain faktorial guna mengeksplorasi kondisi di mana kontribusi AI bersifat aditif atau sinergis terhadap desain pedagogis yang ada.

4. KESIMPULAN

Penelitian ini membuktikan bahwa integrasi chatbot Sokratik berbasis AI ke dalam model inkuiri 5E secara signifikan meningkatkan penalaran ilmiah siswa kelas 5 sekolah dasar ($F(1,57) = 28,45$; $p < 0,001$; $\eta^2 = 0,333$), dengan 74% episode dialog menghasilkan penjelasan berbasis bukti secara mandiri dan penurunan miskonsepsi persisten sebesar 46%. Temuan ini menjawab tujuan penelitian bahwa chatbot Sokratik berbasis aturan mampu berfungsi sebagai mediator epistemis yang memperkuat koordinasi teori-bukti tanpa memberikan jawaban langsung, sekaligus menawarkan model scaffolding individual yang skalabel bagi kelas besar dengan keterbatasan sumber daya guru. Penelitian lanjutan disarankan menggunakan desain faktorial dengan sampel lintas sekolah, periode intervensi yang lebih panjang, dan pengukuran efek keberlanjutan (fading effect) setelah scaffolding AI dihentikan.

DAFTAR PUSTAKA

- [1] OECD, "PISA 2022 Results (Volume I): The State of Learning and Equity in Education," OECD Publishing, Paris, 2023, doi: 10.1787/53f23881-en.
- [2] D. Vo and G. Simmie, "Assessing Scientific Inquiry: A Systematic Literature Review of Tasks, Tools and Techniques," *Int. J. Sci. Math. Educ.*, vol. 23, pp. 871–906, 2024, doi: 10.1007/s10763-024-10498-8.
- [3] S. Afroogh, A. Akbari, E. Malone, M. Kargar, and H. Alambeigi, "Trust in AI: progress, challenges, and future directions," *Humanit. Soc. Sci. Commun.*, vol. 11, 2024, doi: 10.1057/s41599-024-04044-8.
- [4] O. Yetişensoy and H. Karaduman, "The effect of AI-powered chatbots in social studies education," *Educ. Inf. Technol.*, vol. 29, pp. 17035–17069, 2024, doi: 10.1007/s10639-024-12485-6.
- [5] L. Caño and J. Sanz, "Young Children's Epistemic Operations and Emergent Argumentation in Inquiry-Based Discussions: The Role of Teacher Questioning and Collaborative Interactions," *J. Res. Sci. Teach.*, 2025, doi: 10.1002/tea.70010.
- [6] J. Dempere, K. Modugu, A. Hesham, and L. K. Ramasamy, "The impact of ChatGPT on higher education," *Front. Educ.*, vol. 8, 2023, doi: 10.3389/educ.2023.1206936.
- [7] O. De Los Santos et al., "Scientific Thinking Promotes the Development of Critical Thinking in Primary Education," *Educ. Sci.*, 2025, doi: 10.3390/educsci15091174.
- [8] G. Bhusari, K. Tawalare, M. Nikumbh, and K. Tawalare, "Development And Psychometric Validation of Questionnaire to Assess Medovaha Srotas Dushti," vol. 11, no. 22, pp. 4929–4938, 2025.
- [9] P.-B. Degen and I. Asanov, "Beyond Automation: Socratic AI, Epistemic Agency, and the Implications of the Emergence of Orchestrated Multi-Agent Learning Architectures," *ArXiv*, abs/2508.05116, 2025, doi: 10.48550/arxiv.2508.05116.
- [10] T.-T. Wu et al., "Integrating Computational Thinking Diagnostic Mechanism and Reflective Learning: An Innovative Approach to Enhance Learning Outcomes in Introductory Programming," *J. Comput. Assist. Learn.*, vol. 41, no. 5, e70121, 2025, doi: 10.1111/jcal.70121.
- [11] Z. Arkan and M. Bal, "The relationship between school-age students' literacy skills and digital well-being: a systematic review," *BMC Psychol.*, vol. 13, 2025, doi: 10.1186/s40359-025-03573-4.
- [12] H. Fakour and M. Imani, "Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills," *Front. Educ.*, 2025.
- [13] W. Jung, K. Lee, H.-H. Kim, and C. Lim, "How to construct analysis of covariance in clinical trials: ANCOVA with one covariate in a completely randomized design structure," *Korean J. Anesthesiol.*, vol. 78, pp. 315–320, 2025, doi: 10.4097/kja.24820.
- [14] M. Küçükaydin and E. Ayaz, "Validation of the Scientific Reasoning Competencies Instrument: Relationships with Epistemological Beliefs and Analytical Thinking," *Int. J. Sci. Math. Educ.*, vol. 23, pp.

- 343–363, 2024, doi: 10.1007/s10639-024-12485-6.
- [15] J. A. S. Muñoz et al., "GenAI as a cognitive mediator: a critical-constructivist inquiry into computational thinking in pre-university education," *Front. Educ.*, 2025.
- [16] F. Paillusson and M. Booth, "Embracing Representational Plurality to Bypass Misconceptions in Science Education," *Sci. Educ.*, vol. 34, pp. 1955–1969, 2024, doi: 10.1007/s11191-024-00590-4.
- [17] S. Noroozi and H. Karami, "A Rasch-based validation of the University of Tehran English Proficiency Test (UTEPT)," *Lang. Test. Asia*, vol. 14, pp. 1–18, 2024, doi: 10.1186/s40468-024-00290-4.
- [18] Y. Topkaya, Y. Doğan, V. Batdi, and S. Aydin, "Artificial Intelligence Applications in Primary Education: A Quantitatively Complemented Mixed-Meta-Method Study," *Sustainability*, 2025.
- [19] M. Gordon et al., "A scoping review of artificial intelligence in medical education: BEME Guide No. 84," *Med. Teach.*, vol. 46, pp. 446–470, 2024, doi: 10.1080/0142159x.2024.2314198.
- [20] L. Kaldaras et al., "Employing technology-enhanced feedback and scaffolding to support the development of deep science understanding using computer simulations," *Int. J. STEM Educ.*, vol. 11, 2024, doi: 10.1186/s40594-024-00490-7.
- [21] P. M. Newton and S. Jones, "Education and Training Assessment and Artificial Intelligence. A Pragmatic Guide for Educators," *Br. J. Biomed. Sci.*, vol. 81, 2025.
- [22] J. T. E. Richardson, "Eta squared and partial eta squared as measures of effect size in educational research," *Educ. Res. Rev.*, vol. 6, no. 2, pp. 135–147, 2011, doi: 10.1016/j.edurev.2010.12.001.
- [23] L. Wu and K. A. Bakar, "Language Scaffolding Strategies Based on Vygotsky's Zone of Proximal Development in Early Reading Education: A Cross-Cultural Comparison," *Forum Linguist. Stud.*, 2025.
- [24] K. Tang and G. Cooper, "The Role of Materiality in an Era of Generative Artificial Intelligence," *Sci. Educ.*, vol. 34, pp. 731–746, 2024, doi: 10.1007/s11191-024-00508-0.
- [25] I. Pesovski, R. Santos, R. Henriques, and V. Trajkovik, "Generative AI for Customizable Learning Experiences," *Sustainability*, 2024, doi: 10.3390/su16073034.
- [26] Y. Ding, G. Zhu, Q. Bian, and L. Bao, "Analysis of students' conceptual change in learning Newton's third law with an integrated framework of model analysis and knowledge integration," *Phys. Rev. Phys. Educ. Res.*, 2024.
- [27] Y. Hou, Q. Xie, N. Zhang, and J. Lv, "Cognitive load classification of mixed reality human computer interaction tasks based on multimodal sensor signals," *Sci. Rep.*, vol. 15, 2025.
- [28] J. Crawford, M. Cowling, S. Ashton-Hay, J. Kelder, R. Middleton, and G. Wilson, *J. Univ. Teach. Learn. Pract.*, 2023.
- [29] H. Güner and E. Er, "AI in the classroom: Exploring students' interaction with ChatGPT in programming learning," *Educ. Inf. Technol.*, vol. 30, pp. 12681–12707, 2025.
- [30] C. Lin, "Developing Workplace Training and Learning: An Extended Cognitive-Load Framework," *Hum. Resour. Dev. Q.*, 2025, doi: 10.1002/hrdq.21567.
- [31] N. Mathé and A. S. Christensen, "Show and tell: scaffolding practices in lower secondary social science classrooms," *J. Curric. Stud.*, 2024, doi: 10.1080/00220272.2024.2378315.
- [32] M. Nashaat and J. Miller, "Towards Efficient Fine-Tuning of Language Models With Organizational Data for Automated Software Review," *IEEE Trans. Softw. Eng.*, 2024.
- [33] C. Osterhaus and S. Koerber, "The personal epistemology of parents predicts the development of scientific reasoning in children aged 6-10 years," *Dev. Sci.*, 2024.
- [34] X. Lu, Y. Zhang, S. Wu, and G. Shan, "The impact of randomization techniques on the performance of pre-post design models," *BMC Med. Res. Methodol.*, vol. 25, 2025.
- [35] A. Postema et al., "Investigating sources of non-response bias in a population-based seroprevalence study of vaccine-preventable diseases in the Netherlands," *BMC Infect. Dis.*, vol. 24, 2024.
- [36] E. Beard, J. Brown, and L. Shahab, "Sample size requirements to evaluate policies in addiction research using interrupted time series analysis (ITS): Tools and guidance," *Addiction*, 2025.
- [37] C. Xu et al., "Few-shot network intrusion detection method based on multi-domain fusion and cross-attention," *PLOS One*, vol. 20, 2025, doi: 10.1371/journal.pone.0327161.