Pemanfaatan AI dan Hasil Belajar Rasio Siswa Kelas 6 SD/MI: Studi Pra-Eksperimen

Rizki Rahmayani*[1], Sri Nurlailah Saputri[2], Irfan Bahri[3]
[1,2,3] Fakultas Keguruan dan Ilmu Pendidikan, Universitas Tanjungpura
e-mail:*[1]riskrahm@gmail.com, [2]srinurlailasaputri@gmail.com, [3]irfanbahri587@gmail.com

Abstract. *Ratio requires students to understand the relationship between two quantities rather than merely calculate with them, and many sixth graders still find it difficult. Meanwhile, artificial intelligence (AI) is entering classrooms, yet evidence from primary school mathematics, especially on ratio, remains limited. This study examined the difference in ratio learning outcomes between sixth-grade students who learned with a teacher-guided AI tutor and a historical comparison group that learned conventionally. A pre-experimental static-group comparison design was used, in which the comparison group was the sixth-grade cohort of the previous academic year (2025/2026; archived scores, n = 24), not a concurrent control class. The AI group was one sixth-grade class (n = 20) at MI Mas'udi Pontianak in the 2026/2027 academic year, which learned over four meetings with an AI assistant acting as a tutor under the teacher's guidance. Learning outcomes were measured by a 20-item multiple-choice test and analysed with an independent-samples t-test and Cohen's d. The AI group obtained a higher mean ratio score (M = 87.00, SD = 8.18) than the comparison group (M = 66.67, SD = 11.20), t(42) = 6.75, p < .001, d = 2.04. Because there was no pretest, no randomisation, and no concurrent control group, the difference may partly reflect differences between cohorts and cannot be attributed to AI alone; it is an initial indication that needs to be tested in a design with a concurrent control group.*

Keyword: *AI-assisted learning, artificial intelligence, learning outcomes, ratio*

Abstrak. *Rasio menuntut siswa memahami hubungan antarbesaran, bukan sekadar menghitung, dan banyak siswa kelas VI SD/MI masih kesulitan pada materi ini. Di sisi lain, kecerdasan buatan (AI) mulai masuk ke ruang kelas, tetapi buktinya pada jenjang SD/MI, khususnya pada materi rasio, masih terbatas. Penelitian ini bertujuan menguji perbedaan hasil belajar materi rasio antara siswa kelas VI SD/MI yang belajar dengan AI sebagai tutor yang dipandu guru dan kelompok pembanding historis yang belajar secara konvensional. Desain yang dipakai adalah pra-eksperimen static-group comparison dengan kelompok pembanding berupa kelas VI tahun ajaran sebelumnya (2025/2026; data arsip, n = 24), bukan kelas kontrol yang berjalan bersamaan. Kelompok AI adalah satu kelas VI (n = 20) di MI Mas'udi Pontianak tahun ajaran 2026/2027 yang belajar selama empat pertemuan dengan asisten AI sebagai tutor di bawah bimbingan guru. Hasil belajar diukur dengan tes pilihan ganda 20 butir, lalu dianalisis dengan uji-t sampel independen dan Cohen's d. Rerata nilai rasio kelompok AI (87,00) lebih tinggi daripada kelompok pembanding (66,67), t(42) = 6,75; p < 0,001; d = 2,04. Karena tidak ada pretest, pengacakan, maupun kelas kontrol yang berjalan bersamaan, perbedaan ini dapat sebagian berasal dari perbedaan antarangkatan dan tidak dapat dikaitkan dengan AI semata; temuan ini merupakan indikasi awal yang perlu diuji pada desain dengan kelas kontrol yang berjalan bersamaan.*

Kata Kunci: *hasil belajar, kecerdasan buatan, pembelajaran berbantuan AI, rasio*.

PENDAHULUAN

Rasio menuntut siswa memahami hubungan multiplikatif antara dua besaran, bukan sekadar melakukan operasi hitung. Cara berpikir ini menjadi dasar penalaran proporsional yang dipakai pada skala, persen, kecepatan, hingga aljabar (Lamon, 2007), dan materinya diharapkan dikuasai pada akhir fase C Kurikulum Merdeka (Kemendikbudristek, 2022). Namun, kesulitan siswa pada materi ini masih dilaporkan. Kesalahan konseptual yang berulang ditemukan pada siswa kelas VI saat menyelesaikan soal rasio, begitu pula miskonsepsi pada materi perbandingan di tingkat SMP (Fahira et al., 2026). Salah satu bentuknya adalah strategi aditif, yaitu menambahkan bilangan yang sama pada kedua suku sehingga rasio yang senilai dengan 2 : 3 dianggap 4 : 5 ((Lamon, 2007) persoalan serupa juga telah diteliti pada siswa SMP ((Rahaded & Tuasikal, 2025). Data arsip MI Mas'udi Pontianak menunjukkan gambaran yang sejalan: pada materi rasio kelas VI tahun ajaran 2025/2026, hanya 11 dari 24 siswa (45,8%) mencapai kriteria ketuntasan minimal (KKM) 70.

Di sisi lain, kecerdasan buatan (*artificial intelligence*, AI) makin dekat dengan ruang kelas. Berbagai tinjauan menggambarkan AI sebagai tutor, pemberi umpan balik, sekaligus pengatur jalur belajar (Chen et al., 2020; Hwang et al., 2020; Zawacki-Richter et al., 2019), dan AI generatif berbasis percakapan memperluas kemungkinan itu (Baidoo-Anu & Owusu Ansah, 2023). Kajian sistematis cenderung melaporkan dampak positif AI pada pembelajaran matematika (Opesemowo & Adewuyi, 2024), dan bantuan ChatGPT dilaporkan setara dengan bantuan tutor manusia pada keterampilan matematika tertentu (Pardos & Bhandari, 2024). Pada jenjang SD, (Marselina et al., 2026) melaporkan nilai pemecahan masalah matematika yang lebih tinggi pada kelas dengan pembelajaran berbantuan AI daripada kelas kontrol. Namun, gambarannya tidak sepenuhnya cerah: akses bebas ke AI tanpa pengaman dapat membuat siswa SMA tampil lebih buruk begitu bantuan dicabut (Bastani et al., 2025), dan pada beberapa soal penalaran proporsional, alat AI sendiri mereplikasi kesalahan yang lazim dilakukan siswa (Memiş, 2025).

Dari tinjauan tersebut tampak dua celah. Pertama, sebagian besar bukti berasal dari siswa SMA dan mahasiswa; penelitian pada siswa SD/MI masih terbatas, dan sejauh penelusuran penulis belum ada yang menelaah materi rasio secara khusus. Kedua, risiko ketergantungan dan kekeliruan keluaran AI menempatkan guru sebagai penentu cara AI dipakai, padahal kepercayaan diri guru memakai AI belum otomatis berarti kesiapan pedagogis (Julianti et al., 2026). Karena itu, AI dalam penelitian ini dirancang sebagai tutor yang dipandu guru: AI bertanya balik dan memberi petunjuk bertahap, sedangkan guru memeriksa keluaran AI dan memimpin diskusi. Rancangan ini dipilih karena kesalahan seperti strategi aditif berkaitan dengan cara siswa memandang hubungan antarbilangan, sehingga pertanyaan pemantik diharapkan lebih sesuai untuk materi rasio daripada pemberian jawaban langsung.

Kebaruan penelitian ini terletak pada pengujian rancangan tersebut pada materi rasio di kelas VI SD/MI.

Penelitian ini bertujuan menguji perbedaan hasil belajar materi rasio antara siswa kelas VI SD/MI yang belajar dengan AI sebagai tutor yang dipandu guru dan kelompok pembanding historis, yaitu kelas VI tahun ajaran sebelumnya yang belajar secara konvensional. Hipotesis dirumuskan sejalan dengan uji-t dua sisi: H_0: tidak ada perbedaan rerata hasil belajar rasio antara kedua kelompok, dan H_1: terdapat perbedaan rerata hasil belajar rasio antara kedua kelompok.

METODE

Penelitian ini kuantitatif dengan desain pra-eksperimen *static-group comparison* (Sugiyono, 2019). Kelompok pembanding bersifat historis, yaitu nilai asesmen akhir materi rasio kelas VI tahun ajaran 2025/2026 yang belajar tanpa AI, bukan kelas kontrol yang berjalan bersamaan (Tabel 1). Tidak ada *pretest*.

Tabel 1 Desain Penelitian

Kelompok	Perlakuan	Hasil belajar
Kelas VI tahun ini	X	O1
Kelas VI tahun lalu	–	O2

X = pembelajaran dengan pemanfaatan AI; O1 = nilai materi rasio kelas VI tahun ini; O2 = nilai materi rasio kelas VI tahun ajaran sebelumnya (tanpa AI, data arsip).

Penelitian dilaksanakan di MI Mas'udi, Kota Pontianak, semester ganjil tahun ajaran 2026/2027. Kelompok perlakuan adalah satu kelas VI (20 siswa) melalui *total sampling* karena hanya ada satu rombongan belajar; pembandingnya 24 siswa kelas VI tahun ajaran 2025/2026 yang nilai rasionya tersedia pada arsip guru. Nilai tersebut diambil dari arsip nilai materi rasio yang disimpan guru kelas, dengan soal dan cakupan materi yang sama dengan tahun ini. Kondisi pelaksanaan tes tahun lalu tidak dapat diperiksa ulang secara rinci karena hanya data arsip yang tersedia; bagian yang dapat dipastikan dan yang tidak dijelaskan pada paragraf kondisi pelaksanaan tes di bawah. Peneliti sekaligus guru kelas VI pada kedua angkatan. Izin diperoleh dari sekolah dan orang tua, dan data siswa dianonimkan. Siswa memakai AI pada perangkat sekolah dalam kelompok kecil dengan pendampingan guru.

Perlakuan berlangsung empat pertemuan (2 × 35 menit) tentang konsep rasio, rasio ekuivalen, membandingkan rasio, dan soal cerita. Siswa bekerja dalam kelompok kecil dengan ChatGPT dan Gemini secara bersamaan pada perangkat sekolah. AI diberi *prompt* agar bertindak sebagai tutor: bertanya balik, memberi petunjuk bertahap, dan tidak langsung menyebut jawaban akhir sebelum siswa mencoba. Guru memeriksa jawaban AI dan memimpin diskusi di akhir pertemuan. Pada tahun ajaran sebelumnya, materi yang sama diajarkan lewat ceramah, tanya jawab, contoh soal, dan latihan dari buku siswa.

Pembagian penggunaan ChatGPT dan Gemini dalam pembelajaran adalah sebagai berikut. Tiap kelompok memakai satu perangkat dengan ChatGPT dan Gemini dibuka berdampingan, dan kedua aplikasi menerima *prompt* yang sama pada setiap tugas sehingga kelompok dapat membandingkan jawaban keduanya. Penggunaan AI terbatas

pada kegiatan kelompok selama empat pertemuan.

Contoh *prompt* yang dipakai guru untuk mengatur peran AI: “Kamu adalah tutor matematika yang ramah untuk siswa kelas 6 SD/MI. Topik kami adalah rasio. Jangan langsung menyebut jawaban akhir. Mulailah dengan bertanya apa yang sudah kami ketahui, lalu beri petunjuk satu langkah demi satu dan minta kami mencoba lebih dulu. Jika jawaban kami keliru, ajukan pertanyaan yang membantu kami menemukan kesalahannya sendiri. Gunakan bahasa yang sederhana dan contoh dari kehidupan sehari-hari.” Contoh *prompt* yang diketik siswa pada pertemuan rasio ekuivalen: “Kami diminta mencari rasio yang senilai dengan 2 : 3. Kami menjawab 4 : 5 karena menambahkan 2 pada kedua bilangan. Tolong periksa cara berpikir kami tanpa langsung memberi jawaban yang benar.”

Ketika ChatGPT dan Gemini menghasilkan jawaban yang berbeda untuk soal yang sama, guru mengambil langkah berikut. Pertama, kelompok diminta menunjukkan kedua jawaban beserta langkah penyelesaiannya. Kedua, siswa memeriksa kedua jawaban dengan cara yang telah dipelajari, misalnya tabel rasio, penyederhanaan, atau mengembalikan jawaban ke konteks soal. Ketiga, guru menetapkan jawaban yang benar berdasarkan penyelesaiannya sendiri dan kunci jawaban, bukan berdasarkan salah satu aplikasi. Keempat, perbedaan tersebut dibahas dalam diskusi akhir pertemuan, termasuk bila kedua aplikasi sama-sama keliru.

Hasil belajar diukur dengan tes pilihan ganda 20 butir pada lima indikator (Tabel 2), dengan skor 5 untuk jawaban benar dan 0 untuk jawaban salah atau kosong (maksimum 100). Soal dan cakupan materi yang sama dipakai pada kedua angkatan. Validasi isi oleh dosen Tadris Matematika dan kepala madrasah mencakup kesesuaian dengan indikator, ketepatan konsep dan kunci jawaban, kejelasan bahasa, dan kelayakan untuk kelas VI; kesepakatannya dihitung dengan V Aiken(Aiken, 1985), yang sesuai untuk penilaian sejumlah kecil ahli. Prosedur perhitungan, dasar batas validitas, dan rekapitulasi penilaian dijelaskan setelah Tabel 2. Validasi dilakukan setelah data terkumpul sehingga tidak dapat memperbaiki butir sebelum tes diberikan. Ketuntasan mengacu pada KKM madrasah (70).

Tabel 2 Kisi-kisi Tes Hasil Belajar

Indikator	Level kognitif	Butir	Skor maks.
Menyatakan rasio dua besaran dalam konteks sehari-hari	C1–C2	1–4	20
Menyederhanakan rasio dan menentukan rasio yang senilai	C2	5–8	20
Membandingkan dua rasio atau lebih	C2–C4	9–12	20
Menyelesaikan masalah skala (peta/denah)	C3	13–16	20
Menyelesaikan soal cerita perbandingan	C3–C4	17–20	20
Total		**20**	**100**

Validasi isi dilakukan oleh dua validator, yaitu dosen Tadris Matematika dan kepala madrasah. Setiap butir dinilai pada keempat aspek tersebut dengan skala empat pilihan (1 = tidak sesuai, 2 = kurang sesuai, 3 = sesuai, 4 = sangat sesuai). Lembar validasi memakai tanda centang yang dikonversi menjadi skor 4. Keempat aspek dinilai secara

terpisah dan tidak digabung menjadi satu skor per butir. Koefisien V Aiken dihitung untuk tiap aspek pada tiap butir dengan rumus $V = \Sigma s / [n(c - 1)]$, dengan $s = r - l0$, r = skor dari validator, l0 = skor terendah (1), n = banyak validator (2), dan c = banyak kategori skor (4). Jika kedua validator memberi skor 4, maka s = 3 untuk masing-masing validator, $\Sigma s = 6$, dan $V = 6 / (2 \times 3) = 1{,}00$.

Batas $V \geq 0{,}80$ dipakai sebagai kriteria kelayakan butir. Batas itu adalah kriteria konvensional yang lazim dipakai, bukan ambang signifikansi statistik. Dengan dua validator dan empat kategori, V hanya dapat bernilai 0; 0,17; 0,33; 0,50; 0,67; 0,83; atau 1,00, sehingga batas 0,80 pada praktiknya menuntut skor 4 dari kedua validator atau skor 4 dan 3 ($\Sigma s \geq 5$). Bila tiap validator memilih secara acak di antara empat kategori, peluang keduanya sama-sama memberi skor 4 adalah $(1/4)^2 = 0{,}0625$, sehingga V = 1,00 dari dua validator belum dapat dinyatakan bermakna secara statistik pada taraf 0,05 menurut logika uji V Aiken (Aiken, 1985). Karena itu V dalam penelitian ini dibaca sebagai ringkasan kesepakatan yang bersifat deskriptif, bukan bukti statistik validitas. Rekapitulasi penilaian disajikan pada Tabel 3.

Tabel 3 Rekapitulasi Penilaian Validator dan Koefisien V Aiken

Aspek penilaian	Val. 1	Val. 2	Σs	V
Kesesuaian butir dengan indikator	4	4	6	1,00
Ketepatan konsep dan kunci jawaban	4	4	6	1,00
Kejelasan bahasa	4	4	6	1,00
Kelayakan untuk kelas VI	4	4	6	1,00

Val. = validator. Skor berlaku untuk seluruh 20 butir (tanda centang dikonversi menjadi skor 4): kedua validator memberi skor tertinggi (4) pada setiap butir di keempat aspek. s = skor − 1; Σs = jumlah s kedua validator; V = Σs / 6.

Reliabilitas instrumen tidak diuji. Validasi isi oleh dua validator hanya menilai kesesuaian isi, bukan konsistensi pengukuran, sehingga tidak dapat menggantikan pengujian reliabilitas. Reliabilitas juga tidak dapat dihitung secara a posteriori karena hanya skor per indikator yang tersedia, bukan skor per butir.

Kondisi pelaksanaan tes dirangkum sebagai berikut. Kelompok perlakuan mengerjakan tes selama 30 menit, sedangkan durasi pada angkatan sebelumnya tidak tercatat dalam arsip. Tes dikerjakan secara individual dan tertulis. Bantuan yang diperbolehkan adalah alat tulis dan kertas buram, sedangkan bantuan yang tidak diperbolehkan adalah bertanya kepada teman, membuka buku atau catatan, serta memakai kalkulator atau telepon seluler. Siswa kelompok perlakuan mengerjakan tes tanpa bantuan ChatGPT maupun Gemini. Berdasarkan data arsip, hal yang dapat dipastikan untuk angkatan sebelumnya hanyalah skor tiap siswa, butir soal dan kunci jawaban yang sama, serta cakupan materi; durasi, pengawasan, dan bantuan yang diperbolehkan pada angkatan tersebut tidak tercatat dan tidak dapat dipastikan. Ketidakpastian ini membatasi keterbandingan kondisi tes kedua angkatan dan dibahas pada bagian Keterbatasan.

Data dianalisis dengan SPSS pada taraf signifikansi 0,05 (dua sisi). Normalitas diuji dengan Shapiro–Wilk dan homogenitas dengan Levene. Perbedaan rerata nilai rasio diuji dengan uji-t sampel independen karena

kedua kelompok tidak berpasangan, dan ukuran efeknya dihitung dengan Cohen's d berdasarkan simpangan baku gabungan(Cohen, 1988). Perbandingan per indikator memakai uji Mann–Whitney karena skor tiap indikator bersifat diskrit; karena dilakukan lima uji, nilai p tiap indikator yang dilaporkan adalah nilai yang belum disesuaikan, dan signifikansinya ditafsirkan dengan koreksi Bonferroni pada ambang 0,05/5 = 0,01. Kesetaraan kemampuan awal tidak dapat diuji karena tidak ada *pretest* dan siswa tidak diacak.

HASIL DAN PEMBAHASAN

Hasil

Statistik deskriptif nilai rasio kedua kelas disajikan pada Tabel 4. Rerata kelas yang belajar dengan AI (87,00) lebih tinggi daripada kelas tahun lalu (66,67), dengan selisih 20,33 poin dan simpangan baku yang lebih kecil (8,18 dibandingkan 11,20).

Tabel 4 Statistik Deskriptif Hasil Belajar

Kelas	Nilai rasio
Kelas tahun ini (n = 20)	87,00 (8,18)
Kelas tahun lalu (n = 24)	66,67 (11,20)

Angka menunjukkan rerata (simpangan baku).

Uji prasyarat (Tabel 5) menunjukkan data kedua kelas berdistribusi normal (Shapiro–Wilk, p = 0,449 dan p = 0,207) dengan varians homogen (Levene, p = 0,094), sehingga uji-t sampel independen dapat dipakai. Hasil uji-t menunjukkan perbedaan yang signifikan, t(42) = 6,754; p < 0,001, sehingga H_0 ditolak. Ukuran efeknya termasuk kategori besar menurut patokan Cohen (1988), d = 2,04.

Tabel 5 Uji Prasyarat dan Uji Hipotesis

Uji	Statistik	Sig.
Shapiro–Wilk, tahun ini	0,955	0,449
Shapiro–Wilk, tahun lalu	0,945	0,207
Levene, nilai rasio	2,932	0,094
Uji-t, nilai rasio	6,754	< 0,001
Cohen's d	2,04	-

Sebaran nilai menurut kategori disajikan pada Tabel 6. Pada kelas tahun ini, 17 dari 20 siswa (85%) berada pada kategori tinggi, 3 pada kategori sedang, dan tidak ada pada kategori rendah; pada kelas tahun lalu, masing-masing 4, 15, dan 5 dari 24 siswa. Dengan KKM 70, seluruh siswa (20 dari 20; 100%) kelas tahun ini tuntas, dibandingkan 11 dari 24 siswa (45,8%) pada kelas tahun lalu.

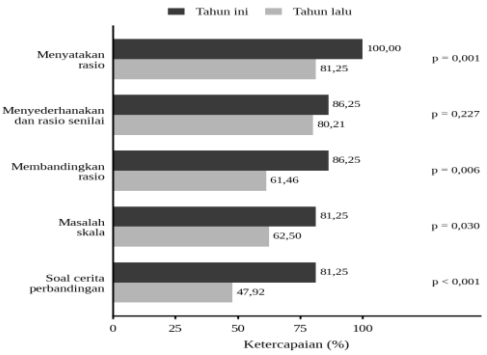
Tabel 6 Sebaran Siswa menurut Kategori Nilai Rasio

Kategori	Tahun ini	Tahun lalu
Tinggi (nilai ≥ 80)	17	4
Sedang (60 ≤ nilai < 80)	3	15
Rendah (nilai < 60)	0	5

Angka menunjukkan banyak siswa.

Rincian per indikator ada pada Gambar 1. Selisih terbesar terlihat pada soal cerita perbandingan, disusul membandingkan rasio. Pada indikator menyatakan rasio seluruh siswa kelas tahun ini mendapat skor penuh, dan selisih terkecil terlihat pada indikator menyederhanakan rasio dan menentukan rasio senilai. Nilai p pada Gambar 1 dan Tabel 7 adalah nilai yang belum disesuaikan untuk perbandingan ganda. Pada α = 0,05 tanpa penyesuaian, empat indikator tampak berbeda signifikan (indikator 1, 3, 4, dan 5). Setelah ambang Bonferroni 0,01 diterapkan pada lima uji, hanya indikator 1 ($p = 0{,}001$), indikator 3 ($p = 0{,}006$), dan indikator 5 ($p < 0{,}001$) yang memenuhi ambang tersebut, sedangkan indikator 4 ($p = 0{,}03$) tidak lagi memenuhinya dan indikator 2 ($p = 0{,}227$) memang tidak signifikan. Dengan demikian,

perbedaan yang dipertahankan setelah penyesuaian hanya pada menyatakan rasio, membandingkan rasio, dan soal cerita perbandingan, dan hasil per indikator tetap dibaca sebagai gambaran eksploratif.



Gambar 1 Ketercapaian Nilai Rasio per Indikator (%)

Ketercapaian = jumlah skor indikator ÷ skor maksimum × 100; p dari uji Mann–Whitney (analisis tambahan; nilai p belum disesuaikan, ambang Bonferroni = 0,01).

Tabel 7 Hasil Uji Mann–Whitney per Indikator dan Interpretasi Ambang Bonferroni

Indikator	p awal	p Bonf. (5p)	Lolos ambang 0,01
Menyatakan rasio	0,001	0,005	Ya
Menyederhanakan rasio	0,227	1,000	Tidak
Membandingkan rasio	0,006	0,030	Ya
Masalah skala	0,030	0,150	Tidak
Soal cerita	< 0,001	< 0,005	Ya

p awal = p belum disesuaikan dari uji Mann–Whitney. p Bonf. = min(1; 5 × p), dihitung dari p yang dibulatkan pada Gambar 1. Keputusan memakai ambang 0,05 / 5 = 0,01.

Pembahasan

Temuan utama penelitian ini adalah rerata nilai rasio kelas yang belajar dengan AI sebagai tutor yang dipandu guru lebih tinggi daripada kelas VI tahun sebelumnya yang belajar secara konvensional, dan perbedaan itu signifikan dengan ukuran efek besar. Data karenanya mendukung H_1. Namun, perbedaan ini perlu dibaca sebagai indikasi adanya perbedaan hasil belajar antarkelompok, bukan sebagai bukti bahwa AI meningkatkan hasil belajar. Kelompok pembanding berasal dari angkatan yang berbeda, siswa tidak diacak, dan kemampuan awal tidak diukur, sehingga selisih 20,33 poin dapat sebagian berasal dari perbedaan kemampuan awal atau kondisi pembelajaran antarangkatan yang tidak terukur.

Arah temuan ini sejalan dengan kajian sistematis yang cenderung melaporkan dampak positif AI pada pembelajaran matematika (Opesemowo & Adewuyi, 2024) dan dengan penelitian pada siswa SD yang melaporkan nilai lebih tinggi pada kelas berbantuan AI (Marselina et al., 2026). Temuan ini juga sejalan dengan laporan bahwa ChatGPT membantu belajar matematika di berbagai jenjang (Wardat et al., 2023), bahwa bantuan AI generatif dapat menyamai bantuan tutor manusia pada keterampilan tertentu (Pardos & Bhandari, 2024), dan bahwa tutor berbasis komputer rata-rata mengungguli pengajaran konvensional(Kulik & Fletcher, 2016). Pada siswa SMA, pengaruh AI terhadap pemahaman konsep matematis (Astriningtyas et al., 2026) serta minat dan hasil belajar (Putri & Haryanto, 2026) juga telah dikaji. Perbedaannya penelitian ini membandingkan dengan angkatan sebelumnya, bukan dengan kelas yang berjalan bersamaan, sehingga kesesuaian arah dengan literatur tidak dapat dibaca sebagai replikasi efek yang sama. Penelitian ini juga tidak membandingkan tutor yang diberi pengaman dengan AI yang diakses bebas seperti pada (Bastani et al., 2025) AI di sini hanya diatur agar tidak langsung memberi jawaban,

sehingga kewaspadaan terhadap risiko ketergantungan tetap diperlukan.

Beberapa mekanisme dapat diajukan sebagai kemungkinan penjelasan, bukan temuan yang telah dibuktikan, karena proses berpikir, jenis kesalahan, interaksi kelompok, dan minat siswa tidak diukur. Pertama, umpan balik datang saat siswa masih berpikir, bukan setelah lembar kerja dikumpulkan. (Bloom, 1984) menunjukkan bahwa bimbingan personal menghasilkan capaian jauh di atas kelas biasa, dan AI memungkinkan pendampingan tiap kelompok kecil yang sulit diberikan guru kepada dua puluh sampai tiga puluh siswa sekaligus. Kedua, AI dapat menyajikan situasi rasio dalam berbagai konteks, misalnya resep, skala peta, dan campuran cat, sehingga siswa dapat melihat bahwa hubungan 2:3 tetap sama walau bendanya berganti. Hal ini berkaitan dengan strategi aditif yang menjadi sumber kesalahan khas, yaitu ketika siswa menambahkan bilangan yang sama pada kedua sisi sehingga memperoleh 4:5 padahal hubungan multiplikatifnya berubah(Lamon, 2007). Dengan *prompt* tutor yang bertanya balik, siswa yang menjawab 4 : 5 dapat diarahkan memeriksa sendiri kekeliruannya lewat tabel rasio atau konteks resep, tetapi karena jenis kesalahan siswa tidak dianalisis, penjelasan ini belum teruji.

Ketiga, secara teoretis pertanyaan balik dan petunjuk bertahap dapat berfungsi sebagai *scaffolding* dalam *zone of proximal development* (Vygotsky, 1978), dan pemecahan soal cerita menjadi langkah-langkah kecil dapat meringankan beban kognitif siswa yang memori kerjanya masih terbatas(Lamon, 2007). Keempat, jawaban AI yang dipakai bersama dalam kelompok dapat menjadi bahan diskusi ketika siswa berbeda pendapat, dan minat belajar yang dilaporkan dipengaruhi AI pada siswa SMA ((Putri & Haryanto, 2026) mungkin ikut berperan. Interaksi kelompok dan minat tidak diukur, dan keterlibatan awal yang tinggi belum tentu bertahan setelah AI tidak lagi terasa baru (efek kebaruan), sehingga keduanya tetap berupa dugaan.

Pola per indikator dan sebaran nilai dibaca secara terbatas. Selisih terbesar pada soal cerita perbandingan dan membandingkan rasio dapat dikaitkan dengan tuntutan kedua indikator itu yang paling banyak berupa penjelasan alasan dan perangkaian beberapa langkah (C3–C4), sedangkan selisih yang tidak signifikan pada menyederhanakan rasio (C2) mungkin karena indikator ini lebih prosedural dan kelas tahun lalu sudah cukup baik lewat latihan berulang. Pada indikator menyatakan rasio, kelas tahun ini mencapai skor penuh sehingga perbedaan lebih lanjut tidak dapat terlihat (efek langit-langit), sedangkan pada soal skala (C3) selisihnya lebih sedang ($p = 0{,}030$) dan tidak lagi memenuhi ambang Bonferroni 0,01, sehingga bersama menyederhanakan rasio ($p = 0{,}227$) indikator ini tidak dapat dinyatakan berbeda secara signifikan setelah penyesuaian. Hanya indikator 1, 3, dan 5 yang tetap signifikan, dan simpulan dibaca selaras dengan ketiga indikator tersebut. Karena nilai p belum disesuaikan, tiap indikator hanya diwakili empat butir, dan analisis butir tidak dilakukan,

pola ini hanya gambaran eksploratif dan tidak dipakai untuk menyimpulkan indikator mana yang paling dipengaruhi AI.

Simpangan baku yang lebih kecil pada kelas tahun ini (8,18 dibandingkan 11,20) menandakan sebaran yang lebih rapat, tetapi uji Levene tidak menunjukkan perbedaan varians yang signifikan, sehingga perbedaan sebaran ini tidak ditafsirkan lebih jauh. Sebaran pada Tabel 6 menunjukkan bahwa selisih rerata tidak hanya ditopang siswa berkemampuan tinggi, karena tidak ada siswa kelas ini pada kategori rendah dan seluruhnya tuntas. Namun, karena angkatannya berbeda, tidak dapat dipastikan apakah siswa yang lemah di kelas ini sebenarnya sudah lebih kuat sejak awal. Pertanyaan itu baru dapat dijawab dengan data kemampuan awal.

Ketiadaan *pretest* membuat penelitian ini tidak dapat menunjukkan besarnya peningkatan yang dialami siswa, dan nilai yang lebih tinggi bisa saja mencerminkan kemampuan awal angkatan yang lebih baik. (Hake, 1998) menekankan perbandingan berbasis peningkatan, yang pada penelitian berikutnya dapat dilakukan lewat *pretest*, *posttest*, dan *N-gain*. Ukuran efek yang besar (d = 2,04) juga perlu dimaknai dengan hati-hati. Pada penelitian jangka pendek dengan sampel kecil, efek cenderung lebih besar daripada yang bertahan ketika perlakuan diperluas ke banyak sekolah, dan besarnya jauh melampaui efek khas sistem tutor cerdas yang dirangkum (Kulik & Fletcher, 2016). Karena itu, hasil ini lebih tepat dipandang sebagai indikasi awal yang menjanjikan daripada estimasi akhir, dan replikasi di beberapa sekolah dengan latar sosial dan ketersediaan perangkat yang berbeda diperlukan sebelum kesimpulan yang lebih luas ditarik.

Perlu dicatat pula bahwa "AI" mencakup banyak jenis alat. (Hanan & Sugiman, 2025) menelaah platform seperti Photomath, Desmos, dan GeoGebra yang berfungsi mengerjakan atau memvisualisasikan soal, sedangkan penelitian ini memakai asisten percakapan yang diarahkan menjadi tutor. Rancangan yang berbeda mungkin menentukan apakah AI menjadi penopang belajar atau sekadar jalan pintas, tetapi penelitian ini tidak membandingkan rancangan yang berbeda. ChatGPT dan Gemini juga dipakai bersamaan sehingga sumbangan masing-masing tidak dapat dipisahkan, dan keluaran kedua alat bisa berbeda untuk pertanyaan yang sama. Penelitian berikutnya dapat memakai satu alat dengan *prompt* yang sama agar perlakuan lebih mudah direplikasi.

Implikasi yang dapat ditarik dari temuan penelitian ini terbatas. Skenario pembelajaran dengan AI sebagai tutor yang dipandu guru dapat dijalankan dengan satu perangkat untuk tiap kelompok, yang lebih realistis bagi sekolah dengan perangkat terbatas, dan kelas yang memakainya mencapai nilai rasio lebih tinggi daripada angkatan sebelumnya, sehingga skenario ini layak diuji lebih lanjut dengan desain yang lebih kuat. Tiga unsur rancangan dapat dicoba guru: meminta AI bertanya balik dan memberi petunjuk bertahap

sebelum menyebut jawaban, membiarkan siswa mencoba lebih dulu dan memakai AI untuk memeriksa pemikiran mereka, serta menutup tiap pertemuan dengan diskusi yang dipimpin guru.

Rekomendasi berikut bersumber dari literatur, bukan dari data penelitian ini. Menyusun *prompt* tutor, menilai keluaran AI, dan memutuskan kapan siswa perlu berhenti bertanya lalu mulai berpikir adalah keputusan pedagogis, dan rasa percaya diri guru memakai AI belum otomatis berarti kesiapan pedagogis (Julianti et al., 2026), sehingga pelatihan perlu berfokus pada rancangan pembelajaran, bukan sekadar pengoperasian aplikasi. Keluaran AI perlu diperiksa karena model bahasa bisa keliru atau terdengar meyakinkan padahal salah (Baidoo-Anu & Owusu Ansah, 2023) termasuk pada soal penalaran proporsional (Memiş, 2025). Akses teknologi dan keahlian pendidik juga masih menjadi tantangan penerapan AI dalam pembelajaran matematika (Marta et al., 2025). Untuk anak SD/MI, penggunaan AI perlu berada dalam pengawasan guru, siswa tidak perlu memasukkan data pribadi, dan sekolah perlu menetapkan aturan yang jelas. Isu bias, privasi, dan keandalan jawaban ini juga dibahas dalam telaah tentang model bahasa besar untuk pendidikan (Kasneci et al., 2023).

Beberapa keterbatasan perlu dicatat. Pertama, kelompok pembanding berasal dari tahun ajaran sebelumnya tanpa pengacakan dan tanpa *pretest*, sehingga perbedaan antarangkatan maupun kondisi pembelajaran tidak dapat disingkirkan. Kedua, peneliti sekaligus guru kelas VI pada kedua angkatan sehingga bias peneliti tidak dapat disingkirkan. Ketiga, perlakuan hanya empat pertemuan dengan sampel kecil sehingga efek kebaruan mungkin ikut berperan, dan ChatGPT serta Gemini dipakai bersamaan sehingga sumbangan masing-masing alat tidak dapat dipisahkan. Keempat, validasi hanya mencakup validitas isi oleh dua validator dan dilakukan setelah data terkumpul, sehingga kesepakatan sempurna pada jumlah validator yang kecil perlu dibaca dengan hati-hati; validitas isi oleh dua validator tidak dapat menggantikan pengujian reliabilitas, dan reliabilitas serta analisis butir tidak dihitung. Akibatnya, besarnya kesalahan pengukuran tidak diketahui sehingga ketepatan interpretasi simpangan baku, hasil uji-t, dan ukuran efek (d = 2,04) tidak dapat dipastikan.

Kelima, tiap indikator hanya diwakili empat butir (skor indikator hanya dapat bernilai 0, 5, 10, 15, atau 20), sehingga cakupan isi terbatas, hasil peka terhadap satu butir, dan terdapat efek langit-langit pada indikator 1; analisis ini juga memakai lima uji dengan nilai p yang belum disesuaikan, sehingga hanya indikator 1, 3, dan 5 yang dipertahankan setelah ambang Bonferroni 0,01. Keenam, kondisi pelaksanaan tes angkatan sebelumnya hanya dapat dipastikan sebatas data arsip, sehingga keterbandingan kondisi tes kedua angkatan tidak sepenuhnya terjamin. Terakhir, retensi jangka panjang, motivasi, dan kemandirian siswa tanpa AI tidak diukur. Penelitian berikutnya dapat menambahkan kelas kontrol yang berjalan bersamaan, *pretest* dan *posttest*, serta *posttest*

tertunda, dan membandingkan beberapa rancangan tutor AI.

SIMPULAN

Terdapat perbedaan hasil belajar rasio yang signifikan antara kelas VI yang belajar dengan AI sebagai tutor yang dipandu guru dan kelas VI tahun ajaran sebelumnya yang belajar secara konvensional: rerata 87,00 berbanding 66,67 ($t(42) = 6{,}754$; $p < 0{,}001$; $d = 2{,}04$), dan seluruh siswa kelas tahun ini (100%) mencapai KKM dibandingkan 45,8% pada kelas tahun lalu. Karena kelompok pembanding berasal dari angkatan berbeda, tanpa pengacakan dan tanpa *pretest*, perbedaan ini tidak dapat ditafsirkan sebagai hubungan sebab-akibat antara AI dan hasil belajar; temuan ini merupakan indikasi awal yang memerlukan pengujian lebih lanjut. Pada analisis per indikator, setelah ambang Bonferroni 0,01 diterapkan, hanya indikator menyatakan rasio, membandingkan rasio, dan soal cerita perbandingan yang berbeda signifikan, sedangkan menyederhanakan rasio dan masalah skala tidak; hasil ini bersifat eksploratif. Guru dapat mencoba skenario tutor AI yang dipandu guru secara terbatas sambil menunggu bukti dari desain yang lebih kuat. Penelitian berikutnya disarankan memakai kelas kontrol yang berjalan bersamaan dan *pretest*, memperluas sampel, serta mengukur retensi belajar.

DAFTAR PUSTAKA

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/00131644854510 12

Astriningtyas, G., Widodo, S. A., & Sulistyowati, F. (2026). Pengaruh teknologi berbasis artificial intelligence terhadap kemampuan pemahaman konsep matematis pada siswa SMA dengan perilaku nomophobia. *JPMI (Jurnal Pembelajaran Matematika Inovatif)*, *9*(3), 471–486. https://doi.org/10.22460/jpmi.v9i3.30856

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, *122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

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

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. https://doi.org/10.1119/1.18809

Hanan, H., & Sugiman, S. (2025). Dampak artificial intelligence terhadap belief peserta didik dalam pembelajaran matematika. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, *9*(1), 339–361. https://doi.org/10.31004/cendekia.v9i1.3852

Julianti, V. M., Rahayu, D. V, & Hermanto, R. (2026). Self-efficacy guru Gen-Z dalam menggunakan artificial intelligence (AI) untuk mendukung pembelajaran matematika. *J-PiMat: Jurnal Pendidikan Matematika*, *8*(1), 2183–2194.

https://doi.org/10.31932/j-pimat.v8i1.6387

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Lamon, S. J. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 629–667). Information Age Publishing.

Marselina, L., Supriadi, & Arisetyawan, A. (2026). Effectivity of artificial intelligence-assisted learning on elementary school students' mathematical problem-solving ability. *Jurnal Pendidikan Profesi Guru*, *4*(2), 1–14. https://doi.org/10.22373/jppg.v4i2.10351

Marta, R., Fadhilaturrahmi, F., & Zulfah. (2025). Artificial intelligence (AI) dalam pembelajaran matematika: Kajian bibliometrik. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, *3*(4), 3236–3245. https://doi.org/10.31004/jerkin.v3i4.1044

Memiş, Y. (2025). Examining the potential and pitfalls of AI in problem solving. *Journal of Pedagogical Research*, *9*(2), 177–190. https://doi.org/10.33902/JPR.202532099

Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(7), em2478. https://doi.org/10.29333/ejmste/14762

Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, *19*(5), e0304013. https://doi.org/10.1371/journal.pone.0304013

Putri, K. P., & Haryanto, H. (2026). Pengaruh penggunaan teknologi berbasis artificial intelligence (ChatGPT) terhadap minat belajar dan hasil belajar matematika siswa SMA swasta Kota Tebing Tinggi. *Ranah Research: Journal of Multidisciplinary Research and Development*, *8*(3). https://doi.org/10.38035/rrj.v8i3.2084

Rahaded, S., & Tuasikal, A. (2025). Karakterisasi penalaran proporsional siswa dalam memecahkan masalah rasio dan proporsi berdasarkan kemampuan dasar matematika. *COGNITIVISM: Journal of Educational Research and Development*, *1*(1), 40–52. https://ejournal.limm-pub.id/index.php/cognitivism/article/view/64

Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272