

Analisis Komparatif Penilaian Diri, Teman, dan Guru dalam Implementasi *Assessment as Learning* pada Materi Sistem Persamaan Linear Dua Variabel di kelas VIII SMP

Suryadi Ishak¹, Ikdafila²

^{1,2}Program Studi Penelitian dan Evaluasi Pendidikan, Program Pascasarjana, Universitas Negeri
Makassar

Email: suryadi.ishak@unm.ac.id

Abstrak

Implementasi Kurikulum Merdeka menuntut siswa memiliki kemampuan metakognisi melalui *Assessment as Learning* (AaL). Penelitian ini bertujuan menganalisis perbandingan persepsi dan asosiasi antara penilaian diri (PD), teman sebaya (PAT), dan guru (PG) pada materi SPLDV di kelas VIII SMP. Metode penelitian kuantitatif dengan desain *within-subject* melibatkan 30 siswa melalui pendekatan longitudinal selama lima pertemuan. Data dianalisis menggunakan *Repeated Measures ANOVA* dan korelasi Pearson. Hasil uji ANOVA menunjukkan tidak terdapat perbedaan rata-rata skor yang signifikan secara kolektif antara ketiga kelompok penilai ($F(2, 58) = 0,465, p = 0,631$), yang mengindikasikan adanya penyelarasan persepsi kelompok. Namun, ditemukan bahwa hipotesis kedua tidak didukung karena korelasi antara penilaian siswa dengan standar guru sangat lemah dan tidak signifikan ($r < 0,3; p > 0,05$). Temuan longitudinal menunjukkan terjadinya konvergensi skor PD dan PAT pada pertemuan kelima, meskipun secara individu siswa masih mengalami *illusion of competence* yang menghambat akurasi penilaian mandiri. Kesimpulannya, meskipun persepsi kelompok telah selaras, kemampuan evaluasi individu siswa masih memerlukan pelatihan kalibrasi metakognitif yang intensif karena rendahnya hubungan linear antara skor siswa dan guru.

Kata Kunci: *Assessment as Learning (AaL), Sistem Persamaan Linear Dua Variabel, Analisis Komparatif.*

PENDAHULUAN

Transformasi pendidikan melalui Kurikulum Merdeka menggeser paradigma evaluasi menuju *Assessment as Learning* (AaL), di mana siswa aktif memantau dan merefleksikan proses berpikir mereka. Dalam matematika, kemampuan ini krusial karena siswa sering kali bergantung pada otoritas guru untuk menentukan kebenaran logis. Khususnya pada materi Sistem Persamaan Linear Dua Variabel (SPLDV), siswa sering mengalami "gegar kognitif" saat bertransisi dari aritmetika konkret ke aljabar abstrak, yang memicu kesalahan prosedural yang fatal (Novita, 2025).

Secara internasional, literasi mengenai pentingnya kemandirian evaluasi siswa telah menjadi tren global. Penilaian mandiri (*self-assessment*) dan penilaian antar teman (*peer-assessment*) bukan sekadar instrumen pengumpulan nilai, melainkan strategi pedagogis untuk membangun *agency* atau kepemilikan belajar pada siswa (Yan dkk., 2023). Ketika siswa mampu menjadi penilai bagi dirinya sendiri, mereka mengembangkan kemampuan untuk mengidentifikasi kesenjangan antara performa saat ini dengan tujuan pembelajaran yang ingin dicapai. Hal ini sejalan dengan konsep "pembelajaran sepanjang hayat" yang ditekankan oleh UNESCO, di mana individu harus memiliki kendali penuh atas navigasi intelektual mereka (Hattie & Clarke, 2018).

Dalam konteks nasional, Kemendikbudristek menekankan bahwa penilaian formatif, khususnya *Assessment as Learning*, merupakan kunci untuk menciptakan pembelajaran yang berdiferensiasi dan bermakna. Namun, urgensi ini seringkali berbenturan dengan realitas di lapangan, terutama pada mata pelajaran yang dianggap sulit seperti matematika. Siswa cenderung memiliki ketergantungan yang tinggi pada otoritas guru dalam menentukan benar atau salah (Oktavihari & Kertiyani, 2025). Oleh karena itu, melatih siswa kelas VIII SMP untuk melakukan penilaian mandiri pada materi abstrak seperti Sistem Persamaan Linear Dua Variabel (SPLDV) menjadi langkah krusial. Tanpa kemampuan *Assessment as Learning*, siswa hanya akan menjadi pengikut prosedur tanpa memahami esensi dari kekeliruan logika yang mereka lakukan.

Implementasi *Assessment as Learning* juga berfungsi sebagai jembatan untuk meningkatkan kejujuran akademik dan objektivitas. Melalui refleksi terstruktur, siswa diajak untuk melihat melampaui angka, berfokus pada kualitas argumen matematis yang mereka susun. Sebagaimana dinyatakan dalam studi terbaru, efektivitas Kurikulum Merdeka sangat bergantung pada sejauh mana ekosistem kelas memberikan ruang bagi siswa untuk berdialog dengan kemampuannya sendiri tanpa rasa takut terhadap hasil skornya (Ardiansyah dkk., 2023).

Beberapa penelitian terdahulu telah berupaya memetakan efektivitas penilaian mandiri dalam pembelajaran matematika. Dalam penelitian (Rusdiana dkk., 2026) menunjukkan bahwa penilaian diri secara kolektif dapat selaras dengan penilaian guru jika didukung oleh rubrik yang transparan. Di sisi lain, (Ma'mun dkk., 2025) menekankan bahwa tanpa umpan balik (*feedback*) yang reguler, penilaian mandiri individu sering kali tidak akurat karena adanya kesenjangan metakognitif antara apa yang dirasakan siswa dengan kualitas nyata pekerjaan mereka. Hal inilah yang menjadi tantangan besar dalam materi Sistem Persamaan Linear Dua Variabel yang menuntut ketelitian prosedural tinggi.

Sistem Persamaan Linear Dua Variabel merupakan salah satu materi dalam kurikulum matematika SMP yang menandai transisi krusial dari berpikir aritmetika konkret ke berpikir aljabar abstrak. Kesulitan utama yang dihadapi siswa terletak pada perubahan paradigma dari memanipulasi angka secara langsung menjadi memanipulasi simbol dan variabel yang merepresentasikan hubungan kuantitatif (Pratiwi & Kurniadi, 2018). Transisi ini seringkali menimbulkan "gegar kognitif" di mana siswa mencoba menerapkan aturan aritmetika pada struktur aljabar yang lebih kompleks, yang berujung pada kesalahan konseptual yang fatal.

Dalam penyelesaian Sistem Persamaan Linear Dua Variabel, metode eliminasi dan substitusi menuntut ketelitian prosedural yang tinggi sekaligus pemahaman logis tentang kesetaraan (*equivalence*). Fenomena yang sering ditemukan di lapangan menunjukkan bahwa banyak siswa merasa telah menyelesaikan soal dengan benar hanya karena mereka berhasil menemukan nilai x dan y , tanpa menyadari bahwa terdapat kesalahan pada langkah manipulasi aljabar atau operasi hitung bilangan bulat di tengah proses (Anisa dkk., 2023). Hal ini menciptakan *false sense of mastery* (perasaan semu telah menguasai), di mana siswa tidak mampu mendeteksi kesalahan prosedural mereka sendiri jika hanya berfokus pada hasil akhir.

Studi internasional menunjukkan bahwa kesalahan dalam aljabar sering kali berakar pada ketidakmampuan siswa dalam melakukan monitoring mandiri selama proses pemecahan masalah berlangsung (Booth dkk., 2014). Pada metode eliminasi, misalnya, siswa seringkali keliru dalam menentukan operasi (penjumlahan atau pengurangan) untuk mengeliminasi variabel, terutama ketika melibatkan koefisien negatif. Tanpa adanya instrumen reflektif seperti *Assessment as Learning*, kesalahan-kesalahan "halus" namun mendasar ini akan terus berulang dan terbawa hingga ke jenjang pendidikan yang lebih tinggi (Cholily dkk., 2025). Oleh karena itu, objektivitas dalam menilai proses bukan hanya sekadar hasil sehingga menjadi kompetensi yang mendesak untuk ditanamkan pada siswa kelas VIII SMP.

Meskipun kajian *Assessment as Learning* (AaL) telah banyak dilakukan, terdapat kesenjangan penelitian pada triangulasi tiga sudut pandang (siswa, teman, guru) secara simultan dalam materi aljabar. Penelitian ini bertujuan untuk mengisi celah tersebut dengan menguji akurasi persepsi siswa melalui pendekatan longitudinal. Berikut rumusan masalah dari penelitian ini:

1. Apakah terdapat perbedaan rata-rata skor antara penilaian diri, teman sebaya, dan guru dalam implementasi AaL pada materi SPLDV?
2. Sejauh mana penilaian diri dan teman sebaya berkorelasi dengan penilaian guru sebagai indikator Hubungan penilaian siswa?

Di Indonesia, dalam konteks Kurikulum Merdeka, penelitian yang secara spesifik membedah reliabilitas penilaian siswa SMP pada materi aljabar masih jarang ditemukan. Kebaruan penelitian ini terletak pada analisis komparatif yang mendalam untuk mengungkap tidak hanya "apa" yang dirasakan siswa tentang kemampuannya, tetapi seberapa akurat perasaan tersebut jika divalidasi oleh rekan sejawat dan guru mata pelajaran.

Hasil penelitian ini diharapkan dapat menjadi rujukan praktis bagi guru dalam merancang rubrik *Assessment as Learning* yang lebih kredibel dan meminimalisir bias penilaian di kelas matematika. Berdasarkan uraian tersebut, hipotesis yang diajukan dalam penelitian ini adalah:

1. Tidak terdapat perbedaan rata-rata skor yang signifikan secara kolektif antara Penilaian Diri (PD), Penilaian Antar Teman (PAT), dan Penilaian Guru (PG) pada materi Sistem Persamaan Linear Dua Variabel.
2. Terdapat hubungan korelasi yang positif dan signifikan antara penilaian siswa (PD dan PAT) dengan penilaian guru (PG) sebagai indikator Hubungan penilaian individu siswa.

METODE

Desain Penelitian

Penelitian ini menggunakan desain komparatif *within-subject*. Desain ini dipilih karena subjek yang sama (30 siswa kelas VIII) dinilai oleh tiga sumber penilai berbeda (diri sendiri, teman sebaya, dan guru) pada variabel yang identik. Data dikumpulkan selama lima pertemuan, di mana skor dari setiap pertemuan dirata-ratakan terlebih dahulu untuk mendapatkan skor komposit sebelum dilakukan analisis komparasi.

Populasi dan Sampel

Subjek dalam penelitian ini adalah siswa kelas VIII di SMP Negeri 1 Sibulue di Kabupaten Bone sebanyak 30 responden. Teknik pengambilan sampel menggunakan *purposive sampling*, di mana responden dipilih karena sedang mempelajari materi Sistem Persamaan Linear Dua Variabel dan telah mendapatkan pembekalan mengenai kriteria penilaian (rubrik) sebelum implementasi *Assessment as Learning* dimulai.

Teknik Analisis Data

Data yang terkumpul diolah secara statistik menggunakan aplikasi JASP (*Jeffreys's Amazing Statistics Program*). Tahapan analisis data meliputi:

1. Statistik Deskriptif: Digunakan untuk melihat nilai rata-rata (*mean*) dan standar deviasi dari masing-masing kelompok penilai.
2. Uji Repeated Measures ANOVA: Digunakan untuk menguji apakah terdapat perbedaan yang signifikan secara statistik antara ketiga kelompok.
3. Uji Post Hoc (Tukey): Dilakukan jika hasil ANOVA menunjukkan adanya perbedaan signifikan, untuk mengetahui pasangan penilai mana yang memiliki perbedaan skor paling tajam.
4. Uji Korelasi Pearson: Digunakan untuk mengukur tingkat akurasi atau keselarasan hubungan antara skor siswa (Diri dan Teman) dengan skor standar dari Guru.

Instrumen Penelitian

Data dalam penelitian ini dikumpulkan menggunakan tiga jenis instrumen penilaian formatif yang dirancang secara paralel, yaitu: Lembar Penilaian Diri (PD), Lembar Penilaian Antar Teman (PAT), dan Lembar Penilaian Guru (PG). Ketiga instrumen tersebut memiliki indikator identik yang diturunkan dari tujuan pembelajaran materi Sistem Persamaan Linear Dua Variabel, sehingga memungkinkan dilakukannya uji komparasi yang valid (*apple-to-apple comparison*). Instrumen ini dikembangkan dalam bentuk skala Likert 5 poin, mulai dari skor 1 (Sangat Kurang/Tidak Pernah) hingga skor 5 (Sangat Baik/Selalu). Secara keseluruhan, terdapat 40 butir pernyataan yang didistribusikan selama lima pertemuan, dengan rincian aspek sebagai berikut:

1. Aspek Pengetahuan (Kognitif): Mengukur pemahaman siswa dalam mengidentifikasi konsep Sistem Persamaan Linear Dua Variabel, membuat model matematika dari masalah kontekstual, serta memilih metode penyelesaian (eliminasi atau substitusi) yang tepat.
2. Aspek Sikap (Afektif): Mengukur karakter siswa selama proses pembelajaran, termasuk kejujuran dalam melaporkan hasil, ketertarikan dalam pembelajaran, dan tanggung jawab.
3. Aspek Keterampilan (Psikomotorik): Mengukur kemampuan teknis siswa dalam melakukan manipulasi aljabar, serta langkah-langkah prosedural penyelesaian sistem persamaan.

Untuk menjamin kualitas data, instrumen ini telah melalui tahap validasi isi oleh ahli (*expert judgment*) untuk memastikan bahwa setiap butir pernyataan selaras dengan kompetensi dasar kelas VIII SMP. Hasil validitas isi menunjukkan bahwa instrumen yang dikembangkan

memiliki koefisien konsistensi internal (δ) sebesar 1 sehingga instrumen layak digunakan. Uji reliabilitas menunjukkan koefisien *Cronbach's Alpha* sebesar 0,78 untuk Penilaian Diri ($\alpha = 0,78$), 0,79 untuk Penilaian Antar Teman ($\alpha = 0,79$), dan 0,80 untuk Penilaian Guru ($\alpha = 0,80$), yang menunjukkan konsistensi internal yang baik. Penggunaan indikator yang sama pada ketiga jenis form bertujuan untuk meminimalkan bias interpretasi, sehingga perbedaan skor yang muncul murni merepresentasikan perbedaan persepsi antara subjek penilai terhadap performa yang sama.

Prosedur Penelitian

Pengumpulan data dalam penelitian ini dilaksanakan melalui tahapan yang sistematis selama lima pertemuan tatap muka pada materi Sistem Persamaan Linear Dua Variabel (SPLDV). Setiap pertemuan dirancang untuk mengintegrasikan kegiatan pembelajaran dengan praktik *Assessment as Learning* (AaL). Prosedur penelitian dibagi menjadi tiga fase utama:

1. Fase Persiapan
2. Fase Implementasi
3. Fase Pengolahan Data

Analisis Data

Data yang telah dikumpulkan melalui lima pertemuan ditabulasi dan dianalisis menggunakan pendekatan statistik kuantitatif dengan bantuan perangkat lunak JASP (*Jeffreys's Amazing Statistics Program*). Analisis data dilakukan dalam tiga tahap utama untuk menguji komparasi dan Hubungan penilaian.

Statistik Deskriptif

Tahap awal analisis melibatkan perhitungan nilai rata-rata (mean), standar deviasi, serta visualisasi data melalui boxplot dan violin plots. Hal ini dilakukan untuk memetakan kecenderungan umum skor yang diberikan oleh ketiga kelompok penilai (PD, PAT, dan PG) serta melihat sebaran konsistensi jawaban responden pada setiap instrumen.

Uji Beda Komparatif (Repeated Measures ANOVA)

Untuk menjawab apakah terdapat perbedaan persepsi yang signifikan antara penilaian diri, antar teman, dan guru, peneliti menggunakan uji Repeated Measures ANOVA. Teknik ini dipilih karena subjek yang dinilai adalah orang yang sama namun diukur oleh tiga sumber penilai yang berbeda (faktor *within-subject*). Asumsi objektivitas dalam *Assessment as Learning* diuji melalui nilai signifikansi (*p - value*) Jika $p < 0.05$, maka terdapat perbedaan persepsi yang nyata antar penilai. Untuk menelusuri pasangan penilai mana yang memiliki perbedaan paling signifikan, dilakukan uji lanjut Post Hoc Comparisons menggunakan koreksi Tukey. Melalui uji ini, peneliti dapat membuktikan secara empiris apakah terjadi fenomena *overestimation* (siswa menilai lebih tinggi dari guru) atau *peer-bias* (teman memberi nilai lebih loyal).

Uji Akurasi (Pearson Correlation)

Meskipun uji ANOVA dapat menunjukkan perbedaan rata-rata, ia tidak dapat menunjukkan apakah penilaian siswa "selaras" secara individu dengan penilaian guru. Oleh karena itu,

peneliti menggunakan *Pearson Correlation Coefficient* (r) untuk mengukur tingkat Hubungan penilaian.

Koefisien korelasi (r) digunakan untuk melihat hubungan antara:

1. Skor Penilaian Diri (PD) dengan Penilaian Guru (PG).
2. Skor Penilaian Antar Teman (PAT) dengan Penilaian Guru (PG).

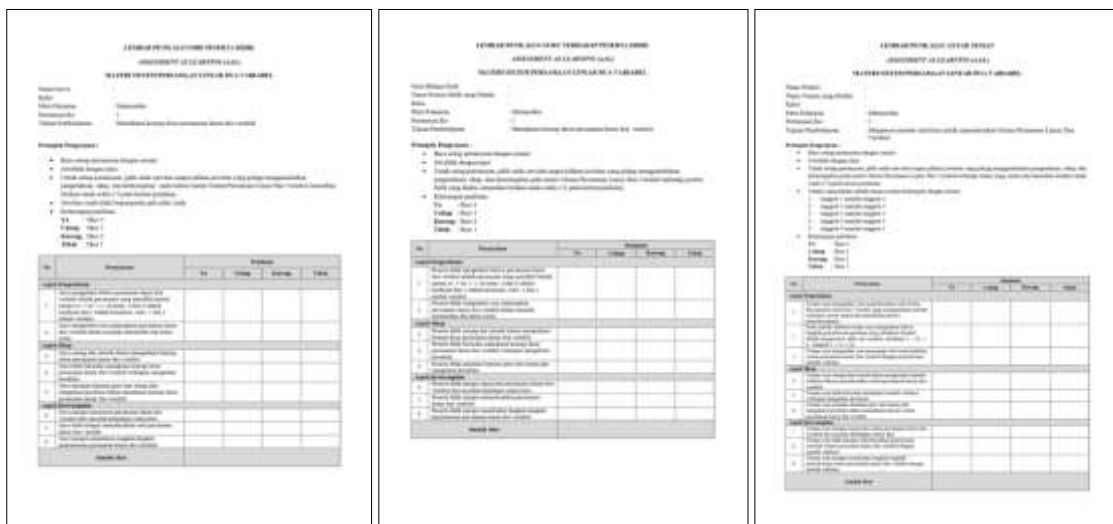
Korelasi yang mendekati $+1$ dengan $p < 0.05$ mengindikasikan bahwa siswa memiliki tingkat akurasi yang tinggi (objektif) dalam mengenali kemampuannya sesuai dengan standar guru. Sebaliknya, korelasi yang rendah atau tidak signifikan menunjukkan bahwa siswa masih mengalami kesulitan dalam menyamakan standar persepsi mereka dengan standar kompetensi Sistem Persamaan Linear Dua Variabel yang ditetapkan oleh guru.

HASIL DAN PEMBAHASAN

Hasil

Fase Persiapan (pra-eksperimen)

Fase persiapan atau pra-eksperimen dalam penelitian ini difokuskan sepenuhnya pada penyesuaian instrumen serta penyamaan persepsi seluruh subjek penelitian. Tahapan ini diawali dengan penyusunan instrumen penelitian yang komprehensif, yakni mengembangkan 40 butir pernyataan yang mencakup tiga aspek utama kompetensi: pengetahuan, sikap, dan keterampilan. Penting untuk dicatat bahwa indikator yang digunakan dibuat identik untuk ketiga jenis formulir penilaian berupa Penilaian Diri (PD), Penilaian Antar Teman (PAT), dan Penilaian Guru (PG) dengan tujuan untuk memastikan objektivitas saat dilakukan uji komparasi nantinya.



Gambar 1. Instrumen *Assessment as Learning* (Penilaian Diri, Penilaian Antar Teman, dan Penilaian Guru)

Setelah draf instrumen selesai disusun, langkah berikutnya adalah melakukan validasi instrumen melalui uji validitas isi (*expert judgment*). Proses ini melibatkan para ahli di bidangnya untuk memastikan bahwa setiap butir pernyataan telah selaras dengan Kompetensi Dasar (KD) materi Sistem Persamaan Linear Dua Variabel pada tingkat kelas VIII SMP serta memenuhi kaidah penulisan instrumen yang baik.

Sebagai langkah penutup dalam fase persiapan, dilakukan sosialisasi rubrik secara mendalam kepada 30 siswa yang menjadi responden. Dalam sesi ini, peneliti menjelaskan tata cara pengisian serta makna esensial dari setiap indikator yang tercantum dalam rubrik. Sosialisasi ini menjadi krusial karena bertujuan untuk meminimalisir bias penilaian yang sering kali muncul akibat ketidaktahuan atau salah persepsi siswa terhadap kriteria penilaian yang ditetapkan, sehingga diharapkan data yang dihasilkan pada fase berikutnya memiliki reliabilitas yang tinggi.

Fase Implementasi

Fase implementasi atau pengumpulan data dilaksanakan secara longitudinal selama lima pertemuan tatap muka yang berfokus pada materi Sistem Persamaan Linear Dua Variabel (SPLDV). Proses ini diawali dengan penyampaian materi oleh guru secara sistematis, mencakup pengenalan konsep SPLDV, penyelesaian menggunakan metode grafik, teknik eliminasi, metode substitusi, hingga pengaplikasiannya dalam pemecahan masalah kontekstual. Struktur pembelajaran ini dirancang untuk memberikan landasan kognitif yang kuat sebelum siswa diminta untuk melakukan evaluasi mandiri.

Inti dari fase ini adalah penerapan praktik *Assessment as Learning* (AaL) yang dilakukan secara konsisten pada setiap akhir pertemuan. Dalam durasi 10 hingga 15 menit terakhir pembelajaran, dilakukan proses pengambilan data serempak yang melibatkan tiga sudut pandang penilai. Pertama, siswa melakukan refleksi terhadap kinerjanya sendiri melalui pengisian formulir Penilaian Diri (PD). Secara bersamaan, setiap siswa juga memberikan penilaian terhadap satu rekan dalam kelompoknya menggunakan formulir Penilaian Antar Teman (PAT). Selama seluruh proses tersebut, guru melakukan observasi intensif dan memberikan penilaian kinerja bagi setiap individu siswa melalui formulir Penilaian Guru (PG).

Data diambil di akhir setiap sesi pembelajaran. Siswa diberikan waktu sekitar 10-15 menit untuk melakukan refleksi. Berikut adalah distribusi fokus materi pada setiap pertemuan:

1. Pertemuan 1: Konsep dasar persamaan linear dua variabel.
2. Pertemuan 2: Konsep Sistem Persamaan Linear Dua Variabel.
3. Pertemuan 3: Penerapan metode eliminasi Sistem Persamaan Linear Dua Variabel.
4. Pertemuan 4: Penerapan metode substitusi Sistem Persamaan Linear Dua Variabel.
5. Pertemuan 5: Penerapan metode kombinasi substitusi-eliminasi untuk menyelesaikan Sistem Persamaan Linear Dua Variabel

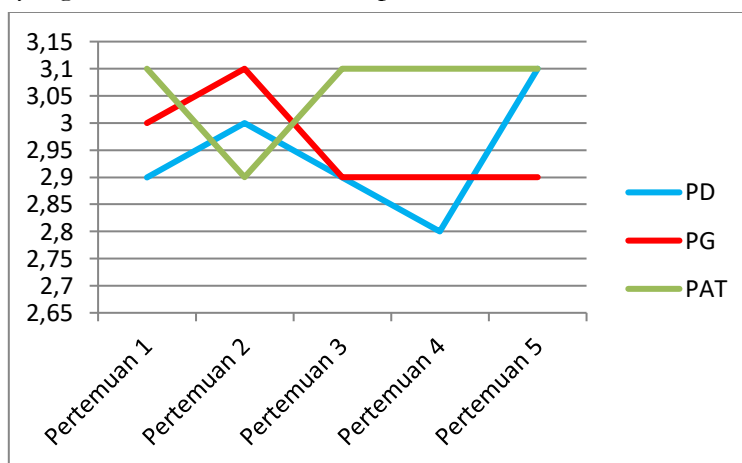
Dengan pendekatan longitudinal ini, data yang terkumpul tidak hanya berupa potret sesaat, melainkan mencerminkan dinamika perubahan standar penilaian siswa terhadap standar penilaian objektif guru di kelas.

Fase Analisis Data

Seluruh data yang terkumpul dari 5 pertemuan (total 40 butir per sumber penilai) ditabulasi ke dalam format Excel, kemudian diimpor ke aplikasi JASP untuk dilakukan analisis mendalam yang meliputi:

Gambar 2. Data skor siswa

Data hasil penilaian selama lima pertemuan pada materi SPLDV menunjukkan dinamika perubahan skor yang bervariasi antar sumber penilai



Gambar 3. Tren Skor Penilaian Diri, Teman, dan Guru (Pertemuan 1-5)

Pada pertemuan pertama, terdapat jarak (gap) yang cukup lebar antara Penilaian Antar Teman ($PAT = 3,1$) dengan Penilaian Diri ($PD = 2,9$), sementara Penilaian Guru ($PG = 3,0$). Memasuki pertemuan kedua hingga keempat, skor ketiga penilai mengalami fluktuasi, di mana skor PD sempat mencapai titik terendah pada pertemuan keempat (2,8). Namun, pada pertemuan kelima, terjadi konvergensi skor di mana PD dan PAT bertemu pada titik yang sama yaitu 3,1, sementara PG stabil di angka 2,9. Secara umum, skor PG menunjukkan tren yang paling stabil dibandingkan PD dan PAT yang lebih fluktuatif sepanjang proses pembelajaran.

Analisis Deskriptif dan ANOVA

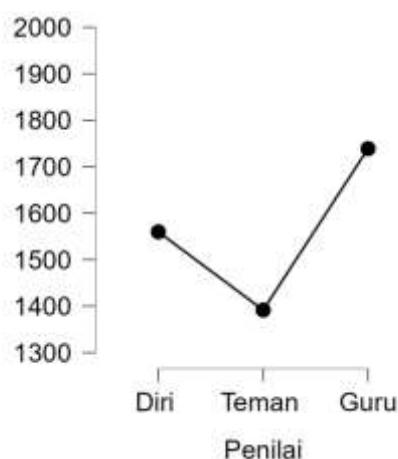
Tahap awal analisis dilakukan dengan membedah karakteristik skor dari ketiga kelompok penilai. Berdasarkan hasil olah data, ditemukan bahwa terdapat variasi rata-rata skor pada implementasi *Assessment as Learning* (AaL) di materi SPLDV. Ringkasan statistik deskriptif disajikan pada Tabel 1 di bawah ini:

Tabel 1. Uji Deskriptif data *Assessment as Learning*

Penilai	N	Mean	SD	SE	Coefficient of variation
Diri	30	1,560	1,495	0,272	0,958
Teman	30	1,392	1,521	0,277	1,093
Guru	30	1,739	1,554	0,283	0,894

Tabel 1 menyajikan gambaran kuantitatif mengenai penilaian dari tiga kelompok (Diri, Teman, dan Guru) terhadap performa siswa pada materi Sistem Persamaan Linear Dua Variabel. Rata-rata (*Mean*) memberikan informasi mengenai kelompok mana yang paling "mudah" atau "ketat" dalam memberikan skor.

1. Guru ($M = 1,739$) : Guru memberikan penilaian dengan rata-rata tertinggi. Hal ini menunjukkan bahwa guru memiliki apresiasi yang lebih besar terhadap proses pengerjaan siswa atau memiliki standar penilaian yang lebih objektif dan mengakomodasi proses logis siswa.
2. Diri ($M = 1,560$) : Siswa menilai diri mereka sendiri lebih rendah dibandingkan penilaian guru. Ini mengindikasikan adanya sikap konservatif atau kurangnya rasa percaya diri siswa terhadap akurasi jawabannya pada materi sistem persamaan linear dua variabel.
3. Teman ($M = 1,392$) : Kelompok teman memberikan rata-rata skor paling rendah. Temuan ini menarik karena mematahkan asumsi umum mengenai *friendship bias* (bias pertemanan); dalam kasus ini, teman justru menjadi penilai yang paling kritis atau ketat di antara ketiga kelompok tersebut.



Gambar 4. *Descriptive Plot* mean data

Variabilitas dan Sebaran Data (Standard Deviation dan Standard Error)

1. *Standard Deviation* (SD): Ketiga kelompok memiliki nilai Standar Deviasi yang relatif berdekatan, berkisar antara 1,495 hingga 1,554. Nilai Standar Deviasi yang cukup besar (mendekati nilai rata-rata) menunjukkan bahwa variasi jawaban antar siswa sangat tinggi, ada siswa yang mendapatkan nilai sangat tinggi dan ada yang sangat rendah secara kontras.

2. *Standard Error (SE)*: Nilai *Standard Error* yang berada di kisaran 0,272 hingga 0,283 mencerminkan tingkat presisi estimasi rata-rata sampel terhadap populasi.

Konsistensi Penilaian (Coefficient of Variation)

Coefficient of Variation (CV) adalah indikator stabilitas atau konsistensi penilaian. Semakin kecil nilainya, semakin konsisten kelompok tersebut dalam memberikan nilai.

1. Penilai Paling Konsisten: Guru ($CV = 0,894$) memiliki nilai koefisien variasi terkecil. Ini membuktikan bahwa guru adalah instrumen penilaian yang paling stabil dan memiliki standar yang jelas dalam memberikan skor.
2. Penilai Paling Tidak Konsisten: Teman ($CV = 1,093$) memiliki variasi penilaian yang paling tinggi. Hal ini menandakan bahwa standar penilaian antar teman masih sangat subjektif dan beragam antar individu (tidak ada keseragaman standar).

Sebelum dilakukan uji *Repeated Measures ANOVA*, dilakukan uji asumsi *sphericity* menggunakan *Mauchly's Test*. Hasil pengujian menunjukkan bahwa asumsi *sphericity* telah terpenuhi ($W = 0,985$; $p > 0,05$), sehingga data layak dianalisis menggunakan *Sphericity Assumed* tanpa perlu koreksi *Greenhouse-Geisser*.

Untuk menguji apakah perbedaan rata-rata tersebut memiliki makna secara statistik, dilakukan uji *Repeated Measures ANOVA* yang disajikan pada Tabel 2:

Tabel 2. Uji *Repeated Measures ANOVA*

<i>Within Subjects Effects</i>					
<i>Cases</i>	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Penilai	$1.809 \times 10^{+6}$	2	904,443	0.465	.631
Residuals	$1.129 \times 10^{+8}$	58	$1.946 \times 10^{+6}$		

Note. Type III Sum of Squares

Hasil pengujian menunjukkan nilai $F(2,58) = 0,465$ dengan nilai signifikansi $p = 0,631$. Karena nilai $p > 0.05$, maka dapat disimpulkan bahwa tidak terdapat perbedaan yang signifikan secara statistik antara ketiga kelompok penilai. Hal ini mengindikasikan bahwa secara kolektif, siswa kelas VIII telah memiliki persepsi yang cukup selaras dengan guru dalam mengimplementasikan *Assessment as Learning (AaL)*.

Analisis Korelasi

Setelah menguji perbedaan rata-rata, langkah selanjutnya adalah menganalisis tingkat akurasi atau keselarasan hubungan antar penilai menggunakan uji *Pearson Correlation*. Analisis ini bertujuan untuk melihat apakah siswa yang memberikan nilai tinggi pada dirinya sendiri juga mendapatkan nilai yang tinggi dari guru, atau sebaliknya. Hasil uji korelasi disajikan pada Tabel 3:

Tabel 3. Uji *Pearson Correlation*

<i>Pearson's Correlations</i>				<i>Pearson's r</i>	<i>P</i>
PD	-	PAT	0.148	.434	
PD	-	PG	0.106	.575	
PAT	-	PG	0.229	.223	

* $p < .05$, ** $p < .01$, *** $p < .001$

Berdasarkan Tabel 3, hasil pengujian menunjukkan fenomena yang menarik terkait Hubungan penilaian mandiri siswa:

1. Hubungan Diri dan Guru (PD-PG): Diperoleh nilai $r = 0.106$ dengan $p = 0.575$. Hal ini menunjukkan masalah pada kalibrasi metakognitif (*metacognitive calibration*). Siswa mengalami *illusion of competence*, di mana mereka merasa menguasai prosedur SPLDV namun gagal mendeteksi kesalahan kecil pada manipulasi aljabar. Temuan bahwa teman sebaya adalah penilai paling kritis (Mean = 1,392) menunjukkan bahwa dalam materi teknis, "mata luar" lebih peka mendeteksi cacat logika pengerjaan dibanding diri sendiri.
2. Hubungan Teman dan Guru (PAT-PG): Diperoleh nilai $r = 0.229$ dengan $p = 0.223$. Meskipun nilai korelasi ini sedikit lebih tinggi dibandingkan penilaian diri, hasilnya tetap tidak signifikan. Ini menunjukkan bahwa teman sebaya pun belum mampu menilai rekannya secara objektif sesuai standar guru.
3. Hubungan Diri dan Teman (PD-PAT): Diperoleh nilai $r = 0.148$ dan tidak signifikan secara statistik $p = 0.434$. Hal ini mengindikasikan bahwa tidak terdapat kesepakatan persepsi yang kuat antara siswa dan teman sejawatnya dalam mengevaluasi capaian pembelajaran. Rendahnya korelasi ini mempertegas bahwa standar penilaian mandiri siswa masih bersifat sangat subjektif dan bervariasi secara individual.

Rendahnya nilai korelasi (r mendekati 0) dan nilai $p > 0.05$ pada seluruh pasangan menunjukkan bahwa meskipun rata-rata kelompok terlihat berdekatan (pada uji ANOVA sebelumnya), namun secara personal/individu, siswa masih mengalami kesulitan dalam melakukan kalibrasi penilaian yang akurat pada materi Sistem Persamaan Linear Dua Variabel.

Pembahasan

Analisis terhadap praktik *Assessment as Learning* (AaL) pada materi Sistem Persamaan Linear Dua Variabel (SPLDV) di kelas VIII mengungkap dinamika menarik antara penilaian mandiri siswa dengan penilaian objektif guru. Untuk tren longitudinalnya mengungkap adanya proses adaptasi metakognitif siswa dalam mengimplementasikan *Assessment as Learning* (AaL). Konvergensi skor PD dan PAT pada pertemuan kelima di angka 3,1 menunjukkan bahwa seiring bertambahnya frekuensi penggunaan rubrik, siswa mulai mencapai stabilitas persepsi kolektif yang serupa dalam menilai performa. Meskipun demikian, skor PG yang

menetap di angka 2,9 pada akhir pertemuan menandakan bahwa guru tetap menjadi penilai yang paling konservatif dan memiliki standar objektif yang tidak mudah berfluktuasi oleh dinamika kelas.

Kenaikan skor PD pada pertemuan terakhir setelah sempat menurun di pertemuan keempat mengindikasikan terjadinya proses *self-correction*. Hal ini berkaitan erat dengan teori kalibrasi metakognitif; melalui latihan penilaian berulang selama lima pertemuan, siswa secara bertahap mampu menyelaraskan estimasi kemampuan mereka dengan indikator yang ada pada rubrik, sehingga meminimalisir fenomena *illusion of competence* yang sering muncul di awal pembelajaran aljabar. Fluktuasi tajam pada skor PAT juga menegaskan bahwa teman sebaya sangat peka terhadap variasi tingkat kesulitan soal SPLDV di setiap pertemuan, yang menjadikan mereka sumber evaluasi yang kritis dalam mendeteksi cacat logika pengerjaan rekan sejawatnya.

Temuan utama penelitian ini menunjukkan bahwa secara kolektif, tidak terdapat perbedaan yang signifikan antara penilaian yang dilakukan oleh Diri (PD), Teman (PAT), dan Guru (PG) dengan nilai $F(2, 58) = 0,465$; $p = 0,631$ pada uji *Repeated Measures ANOVA*.

Tidak ditemukannya perbedaan signifikan menunjukkan bahwa rata-rata skor dari ketiga sumber penilai relatif berdekatan secara kolektif. Namun, temuan ini perlu ditafsirkan secara hati-hati karena tidak selalu berarti penilai sudah selaras secara individu; hal ini bisa dipengaruhi oleh variasi data yang tinggi atau daya uji statistik pada sampel kecil ($N=30$).

Secara pedagogis, hasil ini menunjukkan bahwa fase sosialisasi rubrik dan implementasi *Assessment as Learning* selama lima pertemuan telah berhasil menciptakan "bahasa evaluasi" yang sama di dalam kelas. Siswa tidak lagi memberikan penilaian yang asal-asalan, melainkan sudah mulai mengacu pada indikator yang ditetapkan. Fenomena ini membuktikan bahwa paradigma Kurikulum Merdeka yang mendorong siswa menjadi subjek aktif dalam evaluasi dapat dijalankan dengan baik, bahkan pada materi matematika yang memiliki aturan prosedural kaku seperti Sistem Persamaan Linear Dua Variabel. Ketidaksignifikanan hasil ANOVA mengindikasikan bahwa rata-rata persepsi kelas telah "mengorbit" pada standar yang benar, yaitu standar yang ditetapkan oleh guru.

Salah satu temuan dalam penelitian ini adalah posisi Penilaian Antar Teman (PAT) yang menempati rata-rata skor terendah (1,392) dan divisualisasikan sebagai titik nadir dalam *Descriptive Plot*.

Secara teoritis, penilaian sejawat sering kali dikhawatirkan akan bias karena rasa sungkan atau solidaritas kelompok (*friendship bias*). Namun, data ini menunjukkan kondisi sebaliknya: siswa cenderung lebih kritis dan ketat saat menilai pekerjaan rekannya dibanding saat menilai diri sendiri.

Rendahnya skor PAT ini diduga dipengaruhi oleh karakteristik materi Sistem Persamaan Linear Dua Variabel itu sendiri. Dalam aljabar, kesalahan kecil pada tanda positif/negatif atau kesalahan langkah eliminasi sering kali lebih mudah dideteksi oleh pihak luar (teman) daripada oleh pengerja soal itu sendiri yang mungkin mengalami *blind spot*. Meskipun demikian, tingginya nilai *Coefficient of Variation* pada PAT (1,093) menunjukkan bahwa standar penilaian antar teman masih sangat fluktuatif. Hal ini berarti objektivitas sosial

di tingkat siswa SMP masih belum stabil dan tetap memerlukan pendampingan guru sebagai validator utama.

Meskipun secara rata-rata kelompok tidak ditemukan perbedaan statistik yang signifikan sebagaimana ditunjukkan oleh hasil ANOVA, hasil uji asosiasi mengungkapkan fakta yang kontradiktif pada level individu. Rendahnya korelasi antara PD-PG ($r = 0,106$; $p = 0,575$) dan PAT-PG ($r = 0,229$; $p = 0,223$) menunjukkan bahwa hipotesis kedua dalam penelitian ini tidak terdukung. Hal ini mengindikasikan bahwa secara personal, penilaian siswa belum memiliki hubungan linear yang kuat dengan standar penilaian guru pada materi SPLDV.

Ketidaksinkronan antara persepsi kolektif (ANOVA) dan akurasi individu (korelasi) ini mengarah pada masalah kalibrasi metakognitif (*metacognitive calibration*). Siswa diduga terjebak dalam fenomena *illusion of competence*; mereka merasa telah memahami prosedur umum penyelesaian SPLDV, namun gagal mendeteksi kesalahan-kesalahan logika kecil atau kalkulasi presisi dalam proses pengerjaannya. Rendahnya nilai *mean* pada seluruh kelompok (di bawah 2,0 dari skala 5) mempertegas bahwa rendahnya korelasi bukan hanya masalah bias persepsi, melainkan juga mencerminkan rendahnya capaian kompetensi siswa pada materi aljabar yang bersifat teknis, di mana kesalahan kecil pada satu langkah manipulasi simbol dapat merusak seluruh struktur penyelesaian.

Korelasi yang sangat lemah ini menegaskan bahwa kemiripan nilai rata-rata kelompok (hasil ANOVA) bukanlah representasi dari ketepatan individu, melainkan lebih kepada "kebetulan kolektif" yang belum mencerminkan kemampuan evaluasi individu yang matang. Hal ini berimplikasi pada perlunya bimbingan intensif dari guru. Guru tidak hanya perlu berfokus pada penguasaan materi matematika, tetapi juga pada pelatihan refleksi diri (*self-reflection training*) agar siswa dapat mengidentifikasi letak kekeliruannya secara mandiri dengan lebih akurat lagi.

KESIMPULAN

Berdasarkan analisis data longitudinal selama lima pertemuan pada materi SPLDV, dapat disimpulkan bahwa secara kolektif, tidak terdapat bukti statistik adanya perbedaan rata-rata skor yang signifikan antara Penilaian Diri (PD), Penilaian Antar Teman (PAT), dan Penilaian Guru (PG) ($p = 0,631$). Hal ini mengindikasikan bahwa penggunaan rubrik terstruktur dan sosialisasi yang matang telah membantu menciptakan standar evaluasi kelompok yang relatif seragam. Namun, hasil ini harus ditafsirkan dengan hati-hati mengingat rendahnya perolehan skor rata-rata (di bawah 2,0 dari skala 5), yang menunjukkan bahwa keselarasan ini terjadi pada tingkat capaian kompetensi siswa yang masih rendah.

Ditemukan fenomena unik di mana teman sebaya (PAT) cenderung menjadi penilai yang paling kritis dibandingkan diri sendiri dan guru, yang membuktikan bahwa dalam materi teknis-prosedural, pengamatan eksternal lebih peka dalam mendeteksi cacat logika pengerjaan. Meskipun persepsi kelompok terlihat selaras, hipotesis kedua mengenai adanya korelasi signifikan tidak terdukung. Rendahnya asosiasi individu antara penilaian siswa dengan guru ($r < 0,3$; $p > 0,05$) menegaskan adanya masalah pada kalibrasi metakognitif siswa. Secara personal, siswa masih mengalami *illusion of competence*, di mana mereka mampu memahami indikator umum namun gagal mendeteksi kesalahan kalkulasi

kecil secara presisi. Oleh karena itu, implementasi *Assessment as Learning* di SMP memerlukan pendampingan guru yang lebih intensif untuk mempertajam akurasi evaluasi diri siswa.

DAFTAR PUSTAKA

- Anisa, U. I., Suanto, E., Yuanita, P., & Maimunah, M. (2023). Analisis kesalahan siswa pada pembelajaran matematika berdasarkan teori Newman dalam menyelesaikan soal SPLDV. *Jurnal Paedagogy*, 10(1), 250–258. <https://doi.org/https://doi.org/10.33394/jp.v10i1.6496>
- Ardiansyah, A., Sagita, F., & Juanda, J. (2023). Assesmen dalam kurikulum merdeka belajar. *Jurnal Literasi Dan Pembelajaran Indonesia*, 3(1), 8–13. <https://jurnal.fkip.samawa-university.ac.id/JLPI/article/view/361>
- Booth, J. L., Barbieri, C., Eyer, F., & Paré-Blagojev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *The Journal of Problem Solving*, 7(1), 3. <https://doi.org/10.7771/1932-6246.1161>
- Cholily, Y. M., Ekawati, A. D., Torada, I. Y., Hafizhah, A., Isnaeni, A. F., Yulfa, M. R., Hidayat, W., Wardah, A., Jumaliah, S., & Romadlon, A. K. A. (2025). *Pembelajaran aljabar di sekolah*. UMM Press.
- Hattie, J., & Clarke, S. (2018). *Visible learning: feedback*. Routledge. <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780429485480&type=googlepdf>
- Hidayat, A. (2018). Meta Analisis: Pentingnya Self dan Peer Assesment dalam Pembelajaran. *Jurnal Basicedu*, 2(1), 95–101. <https://www.neliti.com/publications/278075/meta-analisis-pentingnya-self-dan-peer-assesment-dalam-pembelajaran#cite>
- Ma'mun, M., Arsad, M., Jumardiane, L., & Zaenab, S. (2025). Active self assessment sebagai strategi pembelajaran reflektif yang efektif. *At-Tasyrih: Jurnal Pendidikan Dan Hukum Islam*, 11(1), 404–415. <https://doi.org/https://doi.org/10.55849/attasyrih.v11i1.318>
- Novita, I. (2025). *Implementasi Instrumen Assessment as Learning pada Pembelajaran Fisika Berbasis Masalah untuk Menstimulasi Sikap Ilmiah Siswa*. <http://digilib.unila.ac.id/id/eprint/89941>
- Oktavihari, D., & Kertiyani, N. M. I. (2025). Systematic Literature Review: Analisis Hambatan Belajar Siswa dalam Materi Matematika SMP. *Griya Journal of Mathematics Education and Application*, 5(2), 340–354. <https://mathjournal.unram.ac.id/index.php/Griya/index>
- Pratiwi, W. D., & Kurniadi, E. (2018). Transisi kemampuan berpikir aritmatika ke kemampuan berpikir aljabar pada pembelajaran matematika. *Jurnal Gantang*, 3(1), 1–8. <https://doi.org/https://doi.org/10.31629/jg.v3i1.388>
- Rusdiana, A., Imaduddin, A., Chairunnisak, S., Darojah, Z., & Mardiyah, M. (2026). Integrasi perencanaan supervisi dan rubrik penilaian instrumen dalam meningkatkan kualitas pendidikan. *Al-Iryad: Journal of Education Science*, 5(1), 190–214. <https://doi.org/https://doi.org/10.58917/ajjes.v5i1.480>
- Wijayati, I. W. (2025). Urgensi Memahami Ketiganya Bagi Guru dan Mahasiswa Pendidikan. *Asesmen Pembelajaran: Teori Dan Praktik*, 34.

Yan, Z., Panadero, E., Wang, X., & Zhan, Y. (2023). A systematic review on students' perceptions of self-assessment: Usefulness and factors influencing implementation. *Educational Psychology Review*, 35(3), 81.
<https://doi.org/https://doi.org/10.1007/s10648-023-09799-1>