

Profil Kualitas Esai Ilmiah Siswa SMA dalam Kompetisi Biologi: Analisis Skor Per Aspek Penilaian dan Subtema di *Biology Grand Arena* 2026

Ulfa Zahrani^{1*}, Salsabilla Maharani², Nur Faridah³, Amelia Savitri Zairul⁴, Uridia Ulya⁵,
Imroatus Solichah⁶, Bambang Supriatno⁷, Saefuddin⁸, Sariwulan Diana⁹
Program Studi Pendidikan Biologi, FPMIPA, Universitas Pendidikan Indonesia
Jl. Dr. Setiabudhi No. 229, Bandung, Jawa Barat 40154, Indonesia
*e-mail korespondensi: ulfazahrani0411@gmail.com

Abstrak

Penelitian ini memetakan profil kualitas esai ilmiah siswa SMA peserta *Biology Grand Arena* (BGA) 2026 menggunakan rubrik analitik multidimensi, untuk mengidentifikasi aspek penilaian yang paling kuat dan paling lemah sebagai dasar perbaikan pembelajaran biologi SMA. Data diperoleh dari 29 esai yang dinilai oleh tiga juri pada lima aspek yaitu relevansi tema, kedalaman analisis, kekuatan argumentasi, inovasi solusi, serta struktur dan bahasa. Analisis deskriptif menunjukkan bahwa aspek struktur dan bahasa serta relevansi tema memiliki capaian tertinggi (73,4% dan 70,7%), sedangkan kedalaman analisis menempati posisi terendah (57,3%), menandakan kesenjangan kemampuan peserta dalam mengintegrasikan pengetahuan biologis secara mendalam. Korelasi Pearson menunjukkan hubungan signifikan antara seluruh aspek penilaian ($p < 0,01$), dengan korelasi terkuat antara analisis dan argumentasi ($r = 0,899$), yang menegaskan bahwa keterampilan analitis merupakan fondasi bagi argumentasi ilmiah. Hasil uji *One-Way* ANOVA menunjukkan tidak terdapat perbedaan signifikan skor esai antar subtema ($F = 0,514$; $p = 0,726$). Temuan ini memberikan profil empiris kualitas esai ilmiah siswa SMA dalam konteks kompetisi biologi yang dapat menjadi dasar perancangan intervensi pembelajaran yang lebih terarah pada penguatan kemampuan analisis ilmiah di jenjang SMA.

Kata Kunci: Argumentasi Ilmiah, Asesmen Autentik, Esai Ilmiah, HOTS, Pendidikan Biologi.

Abstract

This study maps the quality profile of scientific essays written by high school students participating in Biology Grand Arena (BGA) 2026 using a multidimensional analytic rubric, in order to identify the strongest and weakest assessment aspects as a basis for improving high school biology instruction. Data were obtained from 29 essays evaluated by three judges across five aspects: thematic relevance, depth of analysis, strength of argumentation, solution innovation, and structure and language. Descriptive analysis indicates that the structure-and-language aspect achieved the highest score (73.4%), followed by thematic relevance (70.7%), while depth of analysis was the lowest (57.3%), indicating disparities in participants' ability to integrate biological knowledge. Pearson correlation analysis revealed significant relationships among all scoring aspects ($p < 0.01$), with the strongest correlation between analysis and argumentation ($r = 0.899$), confirming that analytical skill underpins scientific argumentation. One-Way ANOVA results showed no significant difference in essay scores across subtopics ($F = 0.514$; $p = 0.726$). These findings provide an empirical quality profile of high school students' scientific essays in a biology competition context that can inform more targeted instructional interventions to strengthen scientific analysis skills.

Keywords: Authentic Assessment, Biology Education, HOTS, Scientific Argumentation, Scientific Essay.

I. PENDAHULUAN

Pendidikan biologi abad ke-21 menuntut peserta didik tidak hanya menguasai konsep, tetapi juga mampu menginterpretasikan fenomena biologis secara kritis dan menggunakan bukti ilmiah dalam memahami persoalan lingkungan dan kehidupan nyata, seperti perubahan iklim, penurunan kualitas lingkungan, dan hilangnya keanekaragaman hayati. Kemampuan *scientific literacy* dan

higher order thinking skills (HOTS) menjadi kompetensi inti yang perlu dikembangkan secara sistematis, karena peserta didik perlu mampu mengidentifikasi pertanyaan ilmiah, menjelaskan fenomena secara saintifik, dan menggunakan bukti empiris dalam pengambilan keputusan (OECD, 2018). Kompetensi tersebut hanya dapat berkembang melalui pendekatan pembelajaran yang secara eksplisit mendorong penalaran berbasis data dan argumentasi ilmiah, bukan melalui

pembelajaran berbasis hafalan (Fisher & Systems Thinking Association, 2023; Jackson et al., 2023).

Kemampuan argumentasi ilmiah berbasis bukti menjadi bagian penting dalam membantu peserta didik menjelaskan fenomena biologis secara logis, yaitu menyusun klaim, menggunakan data, dan memberikan justifikasi ilmiah (Probosari et al., 2022). Namun, temuan Tanfiziyah dan Rochintaniawati (2021) pada pembelajaran biologi online menunjukkan bahwa argumentasi siswa dalam merespons isu sosiosaintifik masih didominasi oleh klaim (68%) tanpa *warrant* maupun *rebuttal*, mengindikasikan bahwa komponen argumentasi tingkat lanjut masih perlu dilatihkan secara eksplisit. Kemampuan tersebut beririsan langsung dengan keterampilan menulis ilmiah, yang tidak hanya berkaitan dengan kebahasaan, tetapi juga proses mengorganisasikan gagasan, menganalisis masalah, dan mengintegrasikan bukti ke dalam tulisan yang sistematis (Karim, 2023).

Penulisan esai ilmiah merupakan bentuk asesmen autentik yang relevan dalam pendidikan biologi karena mengevaluasi tidak hanya penguasaan konsep, tetapi juga kemampuan peserta didik mengintegrasikan pengetahuan, menyusun argumentasi, dan mengomunikasikan gagasan secara tertulis melalui tugas yang kontekstual dan bermakna (Swaffield, 2011; O'Malley & Pierce, 1996). Esai ilmiah memungkinkan peserta didik mengamati fenomena biologis, menganalisis hubungan sebab-akibat, dan merumuskan solusi berbasis sains secara sistematis (Brown & Abeywickrama, 2010), serta mampu mengungkap kualitas argumentasi dan kemampuan berpikir tingkat tinggi yang sering tidak teridentifikasi melalui tes objektif konvensional (Kuhn, 2019; Osborne, 2010).

Biology Grand Arena (BGA) 2026 merupakan kompetisi biologi tingkat SMA yang menggunakan penulisan esai ilmiah berbasis fenomena biologis kontekstual sebagai instrumen penilaian, dengan rubrik analitik yang memungkinkan evaluasi multidimensional terhadap kualitas berpikir ilmiah peserta (Brown & Abeywickrama, 2010). Studi-studi terdahulu dalam pendidikan biologi sebagian besar masih berfokus pada

pengembangan model pembelajaran atau keterampilan tertentu di ruang kelas, sehingga kajian yang secara khusus memetakan kualitas esai ilmiah siswa SMA dalam konteks kompetisi akademik berbasis rubrik analitik, khususnya berdasarkan aspek penilaian dan subtema yang dipilih peserta secara mandiri masih sangat terbatas di Indonesia.

Penelitian ini berupaya mengisi kekosongan tersebut dengan memetakan kualitas esai ilmiah siswa SMA pada konteks kompetisi biologi nyata (bukan tugas kelas terkontrol) menggunakan rubrik analitik multidimensi, sekaligus menghubungkan profil skor per aspek dengan pola korelasi antar kemampuan argumentasi ilmiah untuk mengidentifikasi aspek yang paling kuat dan paling lemah secara empiris. Pendekatan ini berbeda dari kajian sebelumnya yang umumnya mengevaluasi esai secara holistik atau dalam konteks pembelajaran terstruktur di kelas. Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis profil kualitas esai ilmiah siswa SMA peserta BGA 2026 berdasarkan skor pada setiap aspek penilaian dan subtema, serta mengidentifikasi keterkaitan antar aspek tersebut, sehingga hasilnya dapat menjadi masukan empiris bagi pengembangan pembelajaran biologi yang lebih terarah pada penguatan kemampuan analisis dan argumentasi ilmiah di jenjang SMA.

II. METODE PENELITIAN

A. Desain Penelitian

Penelitian ini menggunakan pendekatan deskriptif-komparatif berbasis analisis dokumen (*document analysis*), karena data bersumber dari dokumen hasil penilaian esai ilmiah sehingga tidak memerlukan intervensi terhadap subjek penelitian (Creswell & Creswell, 2023; Dalglish et al., 2020). Pendekatan deskriptif digunakan untuk memaparkan profil skor esai per aspek penilaian, sedangkan komponen komparatif digunakan untuk membandingkan skor antar subtema menggunakan uji statistik parametrik setelah pengujian asumsi normalitas.

B. Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan bersamaan dengan penyelenggaraan *Biology Grand Arena* (BGA) 2026 di Universitas Pendidikan Indonesia, Bandung. Data esai ilmiah

dikumpulkan pada Januari - Maret 2026, mencakup periode penerimaan naskah, penilaian oleh juri, hingga pengumpulan data skor akhir.

C. Sumber Data

Sumber data dalam penelitian ini adalah naskah esai ilmiah peserta BGA 2026 beserta lembar penilaian digital yang dikelola dalam Google Sheets oleh panitia BGA 2026. Dokumen esai asli juga dikaji secara kualitatif untuk memberikan contoh karakteristik esai berkualitas tinggi dan rendah berdasarkan rubrik, sehingga temuan statistik dapat diinterpretasikan secara kontekstual.

D. Populasi dan Sampel

Populasi sekaligus sampel penelitian adalah seluruh 29 esai ilmiah yang dikumpulkan pada tahap seleksi BGA 2026, ditulis oleh tim peserta SMA dari berbagai daerah di Indonesia yang mendaftar secara sukarela dan melewati tahap administrasi sebelum dinilai. Total sampling digunakan mengingat jumlah esai yang terbatas ($n = 29$) sehingga tidak diperlukan prosedur pengambilan sampel (Creswell & Creswell, 2023). Identitas tim peserta disamarkan dengan nama kode selama proses analisis untuk menjaga kerahasiaan dan objektivitas penilaian.

E. Instrumen Penilaian Esai

Instrumen penilaian yang digunakan adalah rubrik analitik dengan lima aspek berbobot berbeda: (1) relevansi tema (20%), (2) kedalaman analisis ilmiah (25%), (3) kekuatan argumentasi (20%), (4) inovasi solusi (20%), dan (5) struktur dan bahasa penulisan (15%), dengan total skor maksimal 100 poin. Rubrik analitik dipilih karena memungkinkan evaluasi yang lebih transparan dan multidimensional dibandingkan rubrik holistik (Yeo et al., 2024). Setiap esai dinilai secara independen oleh tiga juri berpengalaman di bidang pendidikan biologi, dan skor yang digunakan dalam analisis merupakan rata-rata dari ketiga juri.

F. Validitas dan Reliabilitas Rubrik

Keabsahan dan keterandalan data dijamin melalui beberapa prosedur kontrol kualitas. Pertama, selisih skor total maksimum antar juri untuk setiap esai dihitung sebagai indikator konsistensi antarpemilai, esai dengan selisih ≥ 15 poin dikategorikan memerlukan kalibrasi (*inter-rater calibration*). Hasil rekapitulasi

menunjukkan selisih skor antar juri berkisar antara 4,50 hingga 39,00 poin ($M = 22,02$; $SD = 8,72$), dan sebanyak 23 dari 29 esai (79,3%) memenuhi ambang kalibrasi tersebut. Tingginya proporsi ini menunjukkan variabilitas penilaian antarjuri yang cukup besar dan menjadi catatan penting bagi interpretasi data.

Kedua, perlu diakui secara eksplisit bahwa data yang tersedia berupa skor rata-rata per tim dan selisih skor total antarjuri, tanpa rekaman skor mentah setiap juri pada tiap aspek penilaian. Kondisi ini menyebabkan penghitungan koefisien reliabilitas antarjuri seperti *Intraclass Correlation Coefficient* (ICC) atau *Cronbach's alpha* antarjuri tidak dapat dilakukan secara akurat pada penelitian ini, sehingga selisih skor antar juri tetap digunakan sebagai proksi reliabilitas, sejalan dengan pendekatan yang dilaporkan Yeo et al. (2024), namun dengan penguatan berupa statistik deskriptif selisih skor (rentang, rata-rata, dan SD) di atas. Keterbatasan ini menjadi catatan metodologis yang ditindaklanjuti sebagai rekomendasi bagi penyimpanan data penilaian pada penyelenggaraan kompetisi berikutnya, yakni pencatatan skor mentah setiap juri per aspek agar analisis ICC dapat dilakukan.

Ketiga, triangulasi dilakukan dengan menggabungkan data skor kuantitatif dan kajian kualitatif terhadap naskah esai asli.

G. Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui empat tahap mengacu pada pendekatan READ (*Ready, Extract, Analyse, Distil*) yang dikembangkan Dalglish et al. (2020): (1) persiapan dokumen; (2) ekstraksi skor per aspek per juri ke tabel induk; (3) verifikasi ulang entri skor dan kalkulasi rata-rata; serta (4) analisis data sesuai rencana yang ditetapkan. Tahap verifikasi dilakukan oleh dua anggota tim peneliti secara independen.

H. Teknik Analisis Data

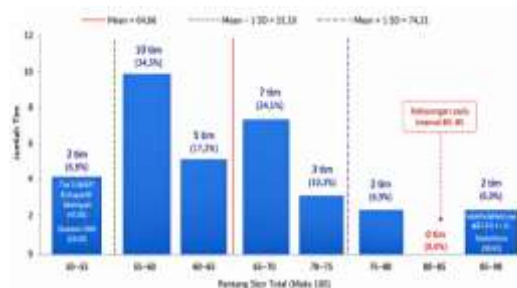
Analisis data dilakukan menggunakan statistik deskriptif dan inferensial parametrik dengan SPSS versi 27 dan Microsoft Excel. (1) Analisis deskriptif digunakan untuk menghitung mean dan SD skor total maupun per aspek; (2) normalitas distribusi diverifikasi menggunakan uji Shapiro-Wilk ($p > 0,05$ menunjukkan distribusi normal); (3) One-Way ANOVA digunakan untuk membandingkan

rata-rata skor total di antara lima subtema (Field, 2018); (4) korelasi Pearson digunakan untuk mengeksplorasi kekuatan dan arah hubungan antar lima aspek penilaian; dan (5) analisis dokumen kualitatif dilakukan untuk mengidentifikasi karakteristik esai berkualitas tinggi dan rendah.

III. HASIL DAN PEMBAHASAN

A. Distribusi Skor Total Esai Peserta

Skor maksimal yang dapat diperoleh setiap esai adalah 100 poin, merupakan akumulasi dari kelima aspek penilaian: Relevansi Tema (20 poin), Kedalaman Analisis (25 poin), Kekuatan Argumentasi (20 poin), Inovasi Solusi (20 poin), dan Struktur serta Bahasa (15 poin). Berdasarkan skor rata-rata 29 tim, diperoleh mean = 64,66; median = 63,67; SD = 9,56; skor minimum 47,00; dan skor maksimum 90,00 (rentang = 43,00). Garis ± 1 SD berada pada rentang 55,10 sampai 74,21.



Gambar 1. Distribusi skor total rata-rata 29 tim peserta *Biology Grand Arena 2026*

Distribusi skor cenderung miring ke kanan (*positively skewed*). Konsentrasi terbanyak berada pada kelas 55 - 60 dengan 10 tim (34,5%), diikuti kelas 65 - 70 dengan 7 tim (24,1%). Sebanyak 20 tim (69,0%) memperoleh skor antara 55 dan 70, menunjukkan bahwa mayoritas peserta berada pada kisaran sedang. Hanya 3 tim (10,3%) yang melampaui ambang mean + 1 SD (> 74,21), yaitu Tesiluriforce (90,00), IndahPadaWaktunya (87,33), dan Biocomplex (77,67), menunjukkan bahwa pencapaian jauh di atas rata-rata relatif langka. Terdapat kekosongan pada interval 80 - 85, menciptakan jurang yang jelas antara Biocomplex (77,67) dan dua tim performa tertinggi (87,33 dan 90,00), yang dapat dipandang sebagai pencilon positif. Sebaliknya, 2 tim (6,9%) berada di bawah ambang mean - 1 SD (< 55,10), yaitu Tim 1 SMAIT Al-Irsyad Al-Islamiyyah (47,00) dan Quantum ABA (54,50). Rentang antar-kuartil

(IQR = 11,33; Q1 = 57,67; Q3 = 69,00) menunjukkan bahwa separuh tim terkonsentrasi pada rentang skor yang relatif sempit, mengindikasikan persaingan ketat pada kelompok tengah.

Tabel 1. Urutan Skor dan Statistik Deskriptif Peserta *Biology Grand Arena 2026*

Peringkat	Nama Tim	Skor Rata-Rata
1	Tesiluriforce	90,00
2	IndahPadaWaktunya	87,33
3	Biocomplex	77,67
4	Algavera	72,33
5	Niners Ngawi	72,33
6	Cleaning serFISH	71,67
7	Inilah Duo Trio	71,00
8	Stom-Amat-A	69,00
9	Oryza Sativa	68,00
10	Genovicta	67,67
11	Monera	67,00
12	Amphibers	66,33
13	StrawVolution	66,00
14	Dream Team	65,67
15	Equaterra	63,67
16	Calon Dokter	63,00
17	Dewa Macaca	61,00
18	Ayam Geprek	59,00
19	Biobestieq Gen Hijau	58,83
20	Tim 2 SMAIT Al Irsyad	58,00
21	Student High	57,83
22	Eco-Logic Squad	57,67
23	Smaywka	57,67
24	Heliocentra	57,67
25	SMP (Sudah Menang Pulang)	56,33
26	Oxygenk	55,50
27	C.B.D Team	55,33
28	Quantum ABA	54,50
29	Tim 1 SMAIT Al-Irsyad	47,00

Statistik Deskriptif	
Mean	64,66
Median	63,37
Std. Deviasi	9,56
Nilai Minimum	47,00
Nilai Maksimum	90,00
Rentang	43,00

Tabel 1 menunjukkan rentang prestasi yang luas di antara 29 tim, dengan selisih 43,00 poin antara skor terendah dan tertinggi. Tesiluriforce menempati posisi puncak (90,00), unggul 2,67 poin dari IndahPadaWaktunya (87,33) dan sekitar 25 poin di atas rata-rata kelompok. Tiga

tim teratas (Tesiluriforce, IndahPadaWaktunya, Biocomplex) berada di atas ambang mean + 1 SD, menunjukkan konsistensi keunggulan lintas aspek penilaian. Sebaliknya, dua tim berada di bawah ambang mean - 1 SD, dengan kelemahan yang nampak terutama pada Kedalaman Analisis dan Inovasi Solusi yang merupakan dua aspek dengan capaian terendah secara keseluruhan (lihat bagian B). Klaster tengah cukup padat, dengan selisih peringkat yang kecil pada beberapa posisi, mencerminkan bahwa kemampuan menghadirkan analisis mendalam dan solusi orisinal menjadi faktor pembeda utama antara tim unggulan dan kelompok tengah, sejalan dengan temuan

Wiggins dan McTighe (2005) bahwa asesmen autentik efektif membedakan kualitas penalaran peserta didik secara multidimensional.

B. Profil Skor Per Aspek Penilaian

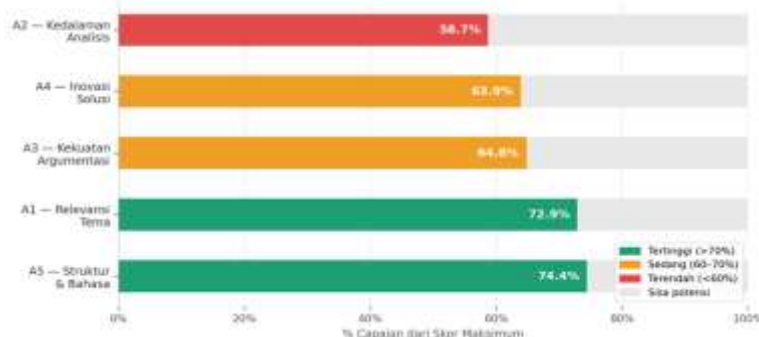
Untuk memahami kekuatan dan kelemahan peserta secara lebih terperinci, dilakukan analisis pada level aspek penilaian. Persentase capaian terhadap skor maksimum masing-masing aspek digunakan sebagai metrik perbandingan yang lebih adil karena mempertimbangkan perbedaan bobot antar aspek.

Tabel 2. Statistik Deskriptif Skor Per Aspek Penilaian Esai

Aspek	Maks	Mean	Median	SD	Min	Maks aktual	% Capaian	Peringkat
Aspek Relevansi Tema	20	14,14	14,00	2,04	10,0	19,0	70,7%	#2
Aspek Kedalaman Analisis	25	14,33	14,00	2,91	10,0	22,0	57,3%	#5 (terendah)
Aspek Kekuatan Argumentasi	20	12,75	12,50	2,02	10,0	18,0	63,7%	#3
Aspek Inovasi Solusi	20	12,42	12,00	2,43	8,0	18,0	62,1%	#4
Aspek Struktur & Bahasa	15	11,02	11,00	1,30	8,33	13,0	73,4%	#1 (tertinggi)

Tabel 2 memperlihatkan kontras menarik antara bobot penilaian dan capaian peserta. Aspek Struktur & Bahasa mencapai persentase capaian tertinggi (73,4%) meskipun memiliki bobot paling kecil (15 poin), dengan SD paling rendah (1,30) yang menggambarkan pemerataan kualitas antar tim. Relevansi Tema menempati posisi kedua (70,7%). Di sisi lain,

Kedalaman Analisis tercatat sebagai aspek dengan capaian terendah (57,3%), walaupun memiliki bobot terbesar (25 poin), dengan SD tertinggi (2,91) yang mengindikasikan variasi kemampuan analitis yang signifikan antar tim. Inovasi Solusi (62,1%) dan Kekuatan Argumentasi (63,7%) menempati posisi tengah.



Gambar 2. Perbandingan % capaian rata-rata per aspek terhadap skor maksimum

Pola capaian ini mengindikasikan bahwa peserta BGA 2026 pada umumnya telah menguasai konvensi penulisan dan pemilihan topik, namun kemampuan untuk melakukan analisis mendalam dan menghasilkan solusi orisinal masih belum merata. Kondisi ini

memiliki implikasi penting bagi pembelajaran biologi yaitu, Kurikulum Merdeka menekankan kompetensi berpikir kritis dan kreatif yang justru tercermin pada aspek Kedalaman Analisis dan Inovasi Solusi, dua aspek capaian terendah dalam penelitian ini. Temuan

ini sejalan dengan hasil Kuhn (2019) yang menunjukkan bahwa kemampuan berpikir kritis sebagai wacana membutuhkan latihan eksplisit dan berkelanjutan.

Temuan serupa dilaporkan oleh Bustami dkk. (2022) dalam studi meta-analisis

C. Perbandingan Skor Antar Subtema

Analisis statistik deskriptif terhadap total skor esai menunjukkan variasi yang bermakna di antara lima subtema yang dipilih. Secara keseluruhan, data skor peserta mencakup rentang nilai antara 47,00 hingga 90,00 dengan rata-rata umum sebesar 64,33 (SD = 9,56), mengindikasikan heterogenitas kemampuan peserta yang cukup luas.

pembelajaran kooperatif, yang menunjukkan bahwa keterampilan berpikir kritis siswa sains cenderung sulit berkembang tanpa model pembelajaran yang secara khusus dirancang untuk melatihnya secara eksplisit.

Tabel 3. Statistik Deskriptif Total Skor Esai per Subtema

Subtema	N	Mean	Median	SD	Min	Maks
Perubahan Mikroklimat	4	60,58	59,92	5,635	54,50	68,00
Keanekaragaman Hayati	6	61,94	60,67	6,075	55,33	71,00
Adaptasi Organisme	1	71,67	71,67	–	71,67	71,67
Mikroorganisme & Lingkungan	3	64,78	59,00	11,182	57,67	77,67
Solusi Biologis Berkelanjutan	15	66,33	66,00	11,423	47,00	90,00

Subtema Adaptasi Organisme memperoleh rata-rata tertinggi (71,67), tetapi hanya diwakili satu peserta (n = 1) sehingga tidak dapat diinterpretasikan sebagai representasi populasi. Keterbatasan ini memengaruhi langsung kekuatan uji beda pada bagian D. Subtema Solusi Biologis Berkelanjutan, kelompok terbesar (n = 15), menunjukkan rata-rata 66,33 (SD = 11,423) dengan rentang skor yang sangat lebar (47,00 - 90,00), mengindikasikan bahwa subtema ini mampu membedakan kemampuan peserta secara lebih nyata, selaras dengan temuan Brookhart (2013) bahwa rubrik yang terbuka terhadap solusi inovatif cenderung menghasilkan varian skor lebih tinggi.

Subtema Mikroorganisme & Lingkungan mencatat rata-rata 64,78 (SD = 11,182; n = 3), dengan selisih mean-median yang mengindikasikan distribusi tidak simetris akibat nilai ekstrem. Subtema Perubahan Mikroklimat memperlihatkan rata-rata terendah (60,58; SD = 5,635; n = 4) dengan rentang lebih sempit, mencerminkan homogenitas kemampuan peserta yang lebih tinggi namun juga keterbatasan eksplorasi dimensi konseptual perubahan iklim mikro. Subtema Keanekaragaman Hayati (n = 6, mean = 61,94) berada di bawah rata-rata keseluruhan, mengisyaratkan bahwa penguasaan argumentasi konservasi berbasis data pada tema ini masih perlu diperkuat (Rustaman, 2011).

D. Hasil One-Way ANOVA

Normalitas distribusi diverifikasi menggunakan uji Shapiro-Wilk, dengan hasil yang mengonfirmasi distribusi normal (p > 0,05), sehingga One-Way ANOVA layak digunakan untuk menguji perbedaan rata-rata skor di antara kelima subtema.

Tabel 4. Hasil *One-Way* ANOVA Perbandingan Skor Esai Antar Subtema

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	201.881	4	50.470	.514	.726
Within Groups	2356.733	24	98.197		
Total	2558.614	28			

Hasil ANOVA menunjukkan $F = 0,514$ dengan $p = 0,726$ ($p > 0,05$), sehingga hipotesis nol yang menyatakan tidak terdapat perbedaan rata-rata skor antar subtema gagal ditolak. Hasil ini perlu dibaca secara hati-hati mengingat ukuran subtema yang sangat tidak seimbang. Subtema Adaptasi Organisme hanya memiliki satu esai ($n = 1$), sehingga kekuatan statistik uji beda ini terbatas dan tidak dapat digunakan untuk menyimpulkan kesetaraan performa antar subtema secara definitif. Dengan keterbatasan tersebut, hasil tidak signifikan ini paling tepat diinterpretasikan sebagai indikasi bahwa kemampuan peserta relatif merata pada

subtema dengan jumlah peserta memadai (Keanekaragaman Hayati, Mikroorganisme & Lingkungan, Perubahan Mikroklimat, Solusi Biologis Berkelanjutan), sementara subtema dengan $n = 1$ sebaiknya dimaknai secara deskriptif saja. Rosenshine (2012) mencatat bahwa ketika peserta memiliki latar belakang keterampilan dasar yang relatif setara, perbedaan topik tidak selalu menghasilkan perbedaan performa yang signifikan, kondisi yang sejalan dengan profil kemampuan dasar peserta BGA 2026 dan *scientific reasoning* secara optimal.

E. Korelasi Antar Aspek Penilaian

Analisis korelasi Pearson dilakukan untuk mengidentifikasi kekuatan dan arah hubungan di antara lima aspek penilaian esai. Tabel 5 menyajikan statistik deskriptif masing-masing aspek, dan Tabel 6 menampilkan matriks korelasi Pearson lengkap.

Tabel 5. Statistik Deskriptif Aspek Penilaian Esai (N = 29)

Statistik	Relevansi (20)	Analisis (25)	Argumentasi (20)	Inovasi (20)	Struktur (15)
Mean	14,14	14,33	12,75	12,42	11,02
Standar Deviasi	2,044	2,911	2,024	2,433	1,302
Minimum	10,00	10,00	10,00	8,00	8,33
Maksimum	19,00	22,00	18,00	18,00	13,00

Tabel 6. Matriks Korelasi Pearson Antar Aspek Penilaian Esai

Aspek	Relevansi	Analisis	Argumentasi	Inovasi	Struktur
Relevansi	1,000	0,822**	0,634**	0,772**	0,540**
Analisis	0,822**	1,000	0,899**	0,818**	0,690**
Argumentasi	0,634**	0,899**	1,000	0,719**	0,672**
Inovasi	0,772**	0,818**	0,719**	1,000	0,539**
Struktur	0,540**	0,690**	0,672**	0,539**	1,000

** Korelasi signifikan pada taraf 0,01 (2-tailed).

Seluruh pasangan aspek menunjukkan korelasi positif yang signifikan pada taraf 99% ($p < 0,01$). Korelasi terkuat ditemukan antara Analisis dan Argumentasi ($r = 0,899$), yang secara empiris menegaskan bahwa kemampuan menganalisis fenomena biologis secara sistematis merupakan prasyarat bagi terbentuknya argumentasi ilmiah yang valid. Pola yang konsisten dengan temuan bahwa Kedalaman Analisis justru menjadi aspek paling lemah peserta (Tabel 2). Hal ini berarti intervensi pedagogis yang menasar kemampuan analisis berpotensi memberi efek

ganda terhadap kualitas argumentasi peserta secara simultan.

Korelasi sangat kuat juga teridentifikasi antara Relevansi dan Analisis ($r = 0,822$) serta antara Analisis dan Inovasi ($r = 0,818$), menunjukkan bahwa peserta yang mampu mengidentifikasi isu relevan cenderung lebih kompeten menganalisisnya secara ilmiah, dan kedalaman analisis berkorelasi positif dengan kualitas solusi inovatif yang diajukan. Korelasi antara Inovasi dan Struktur ($r = 0,539$) serta Relevansi dan Struktur ($r = 0,540$) berada pada kategori sedang, mengisyaratkan bahwa aspek

Struktur relatif lebih independen dari aspek substantif esai lainnya, kemungkinan karena kemampuan struktural lebih dipengaruhi oleh pelatihan teknis penulisan akademik dibandingkan kedalaman literasi saintifik. Pola ini mendukung argumen praktis yaitu,

peningkatan skor Struktur & Bahasa peserta BGA 2026 (capaian tertinggi, Tabel 2) tidak secara otomatis menjamin perbaikan pada kemampuan analisis-argumentasi, sehingga keduanya perlu dilatihkan sebagai kompetensi yang relatif terpisah.

F. Karakteristik Esai Berkualitas Tinggi dan Rendah

Untuk memahami secara lebih konkret apa yang membedakan esai berkualitas tinggi dari esai berkualitas rendah, dilakukan penelaahan pada level butir soal dan indikator penilaian per arena kompetisi. Tabel 7 menyajikan perbandingan karakteristik jawaban berkualitas tinggi dan rendah beserta skor aktual.

Tabel 7. Perbandingan Karakteristik Esai Berkualitas Tinggi dan Rendah Per Arena Kompetisi

Arena	Aspek/Pertanyaan	Jawaban Berkualitas Tinggi	Jawaban Berkualitas Rendah
Arena 1 (Catalyst Q4)	Rantai 3 tahap + angka (maks 8 poin)	Vegetasi Titik B berkurang → cahaya naik dari 800 menjadi ± 40.000 lux & suhu naik 5°C → serangga kehilangan habitat → populasi turun. (8/8; 3 tahap, angka aktual, dampak spesifik)	Jika vegetasi berkurang maka lingkungan berubah dan hewan akan terganggu. (1/8; tanpa angka, tanpa mekanisme)
Arena 2 (Catalyst Q4)	Tanaman bertahan vs berisiko kekeringan (maks 12 poin)	Menyebut nama spesies, ukuran daun aktual, dan mekanisme transpirasi per tanaman. (12/12)	Jawaban umum tanpa nama spesies maupun angka pendukung. (2/12)
Arena 3 (Catalyst Q Homeostasis)	Pola perubahan-respon-tujuan (maks 15 poin)	Aktivitas meningkat → nadi naik dari 72 ke 118 bpm → memenuhi kebutuhan oksigen sel otot (homeostasis). (15/15; 3 bagian pola lengkap + angka aktual)	Tubuh bergerak lalu jantung cepat dan akhirnya kembali normal. (3/15; tanpa angka, tujuan homeostasis tidak dijelaskan)
Arena 4 (Tabel Pengamatan)	Deskripsi Paramecium & Euglena (maks 30 poin)	Menjelaskan bentuk, alat gerak (silia/flagel), dan ciri khusus tiap organisme secara rinci. (28/30)	Hanya menyebut kecepatan gerak dan warna, tanpa menyebut alat gerak. (5/30)
Arena 5 (Catalyst Q4)	Kesimpulan 3 tahap + solusi (maks 8 poin)	Merangkai 3 tahap perubahan dengan 4 data skenario, lalu mengusulkan solusi yang menyentuh akar masalah (<i>root cause</i>). (8/8)	Kesimpulan umum tanpa angka dan tanpa solusi spesifik. (1/8)

Tabel 7 memperlihatkan bahwa perbedaan utama antara esai berkualitas tinggi dan rendah terletak pada kemampuan berpikir mekanistik, kuantitatif, dan sistemik. Esai dengan skor tinggi secara konsisten memuat tiga elemen: (1) nama spesifik atau ilmiah organisme; (2) data kuantitatif dari skenario; dan (3) penjelasan hubungan sebab-akibat yang runtut, ketiganya merupakan komponen yang justru menyusun aspek Kedalaman Analisis, aspek dengan capaian terendah pada Tabel 2. Sebaliknya, esai berskor rendah cenderung umum, minim data, dan tidak menjelaskan mekanisme biologis secara mendalam. Kesenjangan ini paling terlihat pada Arena 2 (Kedalaman Analisis) dan Arena 3 (Kekuatan Argumentasi), yang menunjukkan selisih skor terbesar antara jawaban terbaik dan terburuk. Pola yang konsisten dengan korelasi kuat Analisis-Argumentasi pada bagian E. Temuan ini menegaskan bahwa peningkatan performa peserta memerlukan penguatan latihan penalaran berbasis mekanisme biologis dan

penggunaan data numerik dalam argumentasi, bukan sekadar paparan konten (Osborne, 2010).

G. Implikasi Pembelajaran Esai Ilmiah di SMA

Temuan-temuan di atas menegaskan relevansi pembelajaran berbasis esai ilmiah sebagai strategi pedagogis di jenjang SMA. Sebagai genre teks yang menuntut integrasi pengetahuan, analisis kritis, argumentasi berbasis bukti, dan kreativitas solutif, esai ilmiah mampu mengaktivasi HOTS secara simultan, kompetensi yang secara eksplisit diprioritaskan dalam Kurikulum Merdeka melalui Profil Pelajar Pancasila, khususnya dimensi bernalar kritis dan kreatif. Namun, profil skor penelitian ini menunjukkan bahwa aktivasi tersebut belum optimal pada dimensi analisis, peserta cenderung kuat pada konvensi penulisan (Struktur & Bahasa) tetapi lemah pada penalaran mendalam (Kedalaman Analisis). Kondisi ini sejalan dengan temuan Miharja dan Juniawan (2020) bahwa

kemampuan analisis peserta didik SMA pada materi biologi cenderung rendah apabila tidak didukung oleh bahan atau modul pembelajaran yang secara spesifik dirancang untuk melatih keterampilan tersebut.

Berdasarkan kesenjangan tersebut, implikasi praktis bagi guru biologi SMA terbagi dalam tiga dimensi. Pertama, pada desain pembelajaran, guru perlu mengintegrasikan isu-isu ekologis kontekstual yang menuntut peserta mengurai data kuantitatif dan rantai sebab-akibat secara eksplisit, mengingat justru elemen tersebut yang membedakan esai berkualitas tinggi dan rendah pada Tabel 7. Kedua, pada dimensi pedagogis, guru perlu memfasilitasi latihan analisis teks ilmiah dan diskusi berbasis bukti secara bertahap dengan umpan balik berkelanjutan, karena korelasi kuat Analisis-Argumentasi ($r = 0,899$) menunjukkan bahwa penguatan analisis akan turut mengangkat kualitas argumentasi peserta. Ketiga, pada dimensi asesmen, rubrik analitik lima aspek dalam penelitian ini dapat diadopsi sebagai instrumen asesmen formatif yang memberikan umpan balik spesifik per aspek, bukan sekadar nilai akhir holistik.

Secara lebih luas, kompetisi esai ilmiah berbasis isu biologi seperti BGA berkontribusi pada pengembangan dimensi Profil Pelajar Pancasila, yaitu bernalar kritis melalui argumentasi berbasis data, kreatif melalui solusi inovatif, dan bergotong-royong melalui kerja tim, sehingga integrasi pembelajaran esai ilmiah ke dalam kurikulum biologi SMA relevan secara akademik maupun strategis bagi pembentukan karakter peserta didik.

IV. KESIMPULAN

Penelitian ini memetakan profil kualitas esai ilmiah siswa SMA peserta *Biology Grand Arena* 2026 berdasarkan skor per aspek penilaian dan subtema, dengan tiga temuan utama. Pertama, Struktur & Bahasa merupakan aspek dengan capaian tertinggi (73,4%) dan variasi terendah antar tim, sedangkan Kedalaman Analisis menjadi kelemahan utama peserta dengan capaian terendah (57,3%) dan variasi tertinggi, diperkuat oleh korelasi sangat kuat antara Analisis dan Argumentasi ($r = 0,899$; $p < 0,01$) yang menegaskan kemampuan analitis sebagai fondasi argumentasi ilmiah. Kedua, tidak terdapat perbedaan skor esai yang

signifikan secara statistik antar subtema ($F = 0,514$; $p = 0,726$), meskipun hasil ini perlu dibaca dengan kehati-hatian mengingat ketidakseimbangan ukuran subtema. Ketiga, implikasi pedagogisnya adalah bahwa intervensi pembelajaran biologi SMA perlu diarahkan secara prioritas pada penguatan kemampuan analisis berbasis data dan mekanisme biologis, karena penguatan pada aspek ini berpotensi memberi efek ganda terhadap kualitas argumentasi dan inovasi solusi peserta didik secara simultan. Penelitian lanjutan disarankan menelaah faktor-faktor pengembangan kemampuan analisis dan argumentasi siswa melalui studi longitudinal dan intervensi pedagogis terstruktur, termasuk perbaikan pencatatan data penilaian antarjuri untuk memungkinkan analisis reliabilitas yang lebih kuat (ICC/Cronbach's alpha).

DAFTAR PUSTAKA

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives*. Longman.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices (2nd ed.)*. Pearson Longman.
- Bustami, Y., Mirnawati, & Utami, Y. E. (2022). Model pembelajaran teams games tournament: Studi meta-analisis berpikir kritis dan hasil belajar kognitif siswa sains. *Biosfer: Jurnal Biologi dan Pendidikan Biologi*, 7(1), 30 - 40. <https://doi.org/10.23969/biosfer.v7i1.5454>
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approach (6th ed.)*. SAGE Publications.
- DalGLISH, S. L., Khalid, H., & McMahon, S. A. (2020). Document analysis in health policy research: The READ approach. *Health Policy and Planning*, 35(10), 1424–1431. <https://doi.org/10.1093/heapol/czaa064>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics (5th ed.)*. SAGE Publications.
- Fisher, D. M., & Systems Thinking Association. (2023). Systems thinking activities used in K-12 for up to two decades. *Frontiers in Education*, 8, Article

1059733.
<https://doi.org/10.3389/feduc.2023.1059733>
- Jackson, W. M., Binding, M. K., Grindstaff, K., Hariani, M., & Koo, B. W. (2023). Addressing sustainability in the high school biology classroom through socioscientific issues. *Sustainability*, 15(7), Article 5766. <https://doi.org/10.3390/su15075766>
- Karim, A. R. (2023). Analisis pentingnya kemampuan menulis karya ilmiah pada siswa SMA. *NUSRA: Jurnal Penelitian dan Ilmu Pendidikan*, 4(4), 1226–1233. <https://doi.org/10.55681/nusra.v4i4.1802>
- Kuhn, D. (2019). Critical thinking as discourse. *Human Development*, 62(3), 146–164. <https://doi.org/10.1159/000500171>
- Kuhn, D., & Franklin, S. (2006). *The second decade: What develops (and how)*. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 2: Cognition, perception, and language (6th ed.)*. Wiley.
- Miharja, J., & Juniawan, E. A. (2020). Pengembangan modul integrated multimedia pada materi sistem ekskresi manusia untuk meningkatkan kemampuan analisis peserta didik di SMA. *Biosfer: Jurnal Biologi dan Pendidikan Biologi*, 2(5), 1–9.
- OECD. (2018). *PISA 2018 assessment and analytical framework: Science, reading, mathematics, financial literacy and global competence*. OECD Publishing.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Addison-Wesley.
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328(5977), 463–466. <https://doi.org/10.1126/science.1183944>
- Probosari, R. M., Sajidan, S., Suranto, S., & Prayitno, B. A. (2022). Integrating reading as evidence to enhance argumentation in scientific reading-based inquiry: A design-based research in biology classroom. *Jurnal Pendidikan IPA Indonesia*, 11(1), 171–184. <https://doi.org/10.15294/jpii.v11i1.29350>
- Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator*, 36(1), 12–19.
- Rustaman, N. Y. (2011). *Pendidikan dan penelitian sains dalam mengembangkan keterampilan berpikir tingkat tinggi untuk pembangunan karakter*. Prosiding Seminar Nasional FMIPA UNY.
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433–449. <https://doi.org/10.1080/0969594X.2011.582838>
- Tanfiziyah, R., & Rochintaniawati, D. (2021). Profil kemampuan argumentasi siswa mengenai isu sosiosaintifik dalam pembelajaran online. *Biosfer: Jurnal Biologi dan Pendidikan Biologi*, 6(1), 6–14. <https://doi.org/10.23969/biosfer.v6i1.4081>
- Wiggins, G., & McTighe, J. (2005). *Understanding by design (2nd ed.)*. ASCD.
- Yeo, S. W., Signorelli, C., Vo, K., & Smith, G. (2024). A retrospective cohort analysis comparing analytic and holistic marking rubrics in medical research education. *Journal of Medical Education and Curricular Development*, 11, 1–9. <https://doi.org/10.1177/23821205241277337>