



<http://dx.doi.org/10.25157/jwp.v%vi%i.23388>

Evaluasi Pembelajaran Bahasa dan Sastra Indonesia Berbasis Teori H. Douglas Brown

¹Suhendar, ¹Aditya Ansor Alsunah, ¹Agus Hamdani

¹IPI, Garut, Indonesia

Email: suhendarbiya@gmail.com

Abstract

Evaluation of Indonesian language and literature learning is still often limited to written tests that assess final products, while a framework teachers can use to examine the quality of their instruments has received little attention. This study aims to map H. Douglas Brown's five principles of language assessment (practicality, reliability, validity, authenticity, and washback) onto Indonesian language and literature learning, to develop example instruments based on these principles, and to identify the challenges of applying them. The study used a descriptive qualitative method with a library research design covering 22 sources published between 2016 and 2026: one main book by Brown and Abeywickrama, 18 journal articles, one conference proceeding, and two methodology books. The data were analyzed through data condensation, presentation in synthesis matrices, and conclusion drawing. The results show three findings. First, the five principles can be translated into concrete quality-check questions for assessing language skills and literary appreciation. Second, the five instruments developed, namely literary appreciation presentations, portfolios, performance projects, creative writing, and self-reflection, each stand out on particular principles and have characteristic weaknesses, such as rater reliability in oral tasks and time burden in portfolios. Third, the main challenges of implementation are inconsistent rubrics across raters, limited time and teacher competence, and the interpretive nature of literary works. The novelty of this study lies in using Brown's five principles as a single analytical framework for both language and literature learning in Indonesian.

Keywords: learning evaluation, Indonesian language and literature, H. Douglas Brown, authentic assessment, washback

Abstrak

Evaluasi pembelajaran Bahasa dan Sastra Indonesia masih kerap berhenti pada tes tertulis yang menilai hasil akhir, sementara kerangka yang dapat dipakai guru untuk memeriksa mutu instrumennya belum banyak dibahas. Penelitian ini bertujuan memetakan lima prinsip evaluasi bahasa H. Douglas Brown (kepraktisan, reliabilitas, validitas, autentisitas, dan *washback*) ke dalam pembelajaran Bahasa dan Sastra Indonesia, menyusun contoh instrumen berdasarkan prinsip tersebut, dan mengidentifikasi tantangan penerapannya. Penelitian menggunakan metode kualitatif deskriptif dengan studi pustaka terhadap 22 sumber yang terbit pada 2016–2026, yaitu satu buku utama Brown dan Abeywickrama, 18 artikel jurnal, satu prosiding, dan dua buku metodologi. Data dianalisis melalui kondensasi data, penyajian dalam matriks sintesis, dan penarikan simpulan. Hasil kajian menunjukkan tiga hal. Pertama, kelima prinsip Brown dapat diturunkan menjadi pertanyaan pemeriksa mutu yang konkret untuk menilai keterampilan berbahasa dan apresiasi sastra. Kedua, lima instrumen yang disusun, yaitu presentasi apresiasi sastra, portofolio, proyek pementasan, menulis kreatif, dan refleksi diri, masing-masing menonjol pada prinsip tertentu dan memiliki kelemahan khas, misalnya reliabilitas penilai pada tugas lisan dan beban waktu pada portofolio. Ketiga, tantangan utama penerapannya adalah rubrik yang belum konsisten antarpemilai, keterbatasan waktu dan kompetensi guru, serta sifat penafsiran karya sastra. Kebaruan kajian ini terletak pada pemakaian kelima prinsip Brown sebagai satu kerangka analisis untuk pembelajaran bahasa sekaligus sastra Indonesia.

Kata Kunci: evaluasi pembelajaran, Bahasa dan Sastra Indonesia, H. Douglas Brown, penilaian autentik, *washback*



Karya ini dilisensikan di bawah [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

Cara Sitasi:

Suhendar, et.al. (2026). Evaluasi Pembelajaran Bahasa dan Sastra Indonesia Berbasis Teori H. Douglas Brown. *Jurnal Wahana Pendidikan*. 13(2), 727-738.

Sejarah Artikel:

Dikirim 05-02-2026, Direvisi 08-10-2026, Diterima 17-10-2026.

PENDAHULUAN

Evaluasi pembelajaran tidak hanya berfungsi mencatat hasil belajar. Evaluasi juga memberi tahu guru dan peserta didik bagian mana yang sudah dikuasai dan bagian mana yang masih perlu diperbaiki, sehingga penilaian dipahami sebagai bagian dari proses belajar itu sendiri (Black & Wiliam, 2018). Pandangan ini sejalan dengan arah Kurikulum Merdeka yang menempatkan asesmen sebagai bagian yang menyatu dengan pembelajaran (Wicaksono et al., 2025).

Pada mata pelajaran Bahasa dan Sastra Indonesia, evaluasi menghadapi tantangan yang khas. Yang dinilai bukan hanya pengetahuan tentang kaidah bahasa, tetapi juga keterampilan menyimak, berbicara, membaca, dan menulis, serta kemampuan menafsirkan dan mengapresiasi karya sastra. Aspek yang terakhir bersifat interpretatif, sehingga jawaban peserta didik tidak selalu dapat dinilai dengan kunci benar-salah. Kajian sistematis atas penelitian di Indonesia menunjukkan bahwa praktik evaluasi pembelajaran, termasuk evaluasi menulis, masih banyak berfokus pada hasil akhir dan belum memantau proses belajar (Mulyadi et al., 2026).

Berdasarkan penelitian terdahulu dapat diidentifikasi sedikitnya empat permasalahan. Pertama, evaluasi cenderung berorientasi pada hasil akhir dan kurang menangkap proses (Mulyadi et al., 2026; Wicaksono et al., 2025). Kedua, rubrik penilaian belum seragam antarguru sehingga konsistensi penilaian lemah, sedangkan validitas dan reliabilitas instrumen jarang dilaporkan (Wicaksono et al., 2025). Ketiga, guru masih terkendala pemahaman, pelatihan, dan waktu dalam merancang penilaian autentik (Friyandini, 2025; Wicaksono et al., 2025). Keempat, penilaian pembelajaran sastra yang bersifat interpretatif memerlukan kriteria yang jelas agar tetap adil, dan mutu rubrik sangat bergantung pada ketepatan kriteria tersebut (Brookhart, 2018).

Salah satu kerangka yang dapat dipakai untuk menjawab persoalan itu yaitu lima prinsip evaluasi bahasa yang dirumuskan H. Douglas Brown, yaitu kepraktisan (*practicality*), reliabilitas (*reliability*), validitas (*validity*), autentisitas (*authenticity*), dan dampak balik (*washback*) (Brown & Abeywickrama, 2019). Kelima prinsip ini tidak berdiri sendiri. Validitas, misalnya, menyangkut kelayakan penafsiran dan penggunaan hasil penilaian, dan untuk itu diperlukan bukti, antara lain bahwa skor yang dihasilkan cukup konsisten (Kane, 2016).

Penelitian tentang evaluasi pembelajaran Bahasa Indonesia sebenarnya sudah cukup banyak, namun sebagian besar membahas bentuk asesmen atau keterampilan tertentu. Salah satunya, yaitu pengembangan instrumen penilaian otentik berbasis kecakapan hidup (Sa'idah et al., 2020), pengembangan penilaian autentik pada teks eksposisi (Susilo et al., 2021), persepsi guru

terhadap penilaian autentik untuk keterampilan membaca (Sari & Sukirno, 2021), pelaksanaan penilaian autentik pada masa pandemi (Sunu et al., 2021), dan penilaian autentik bermuatan literasi digital di SMK (Fadilla et al., 2023). Tinjauan yang lebih luas, baik berupa kajian sistematis (Mulyadi et al., 2026; Wicaksono et al., 2025) maupun telaah teoretis (Friyandini, 2025), merangkum praktik dan hambatan penilaian, tetapi tidak memakai prinsip evaluasi bahasa Brown sebagai kerangka analisis yang utuh. Adapun *gap* penelitian ini yaitu penulis belum menemukan kajian yang memakai kelima prinsip Brown secara bersamaan untuk menelaah evaluasi pembelajaran Bahasa dan Sastra Indonesia. Kedua, kajian yang ada lebih banyak menyoroti keterampilan berbahasa, khususnya menulis, daripada apresiasi sastra. Ketiga, pembahasan mutu instrumen, seperti validitas dan reliabilitas, belum diterjemahkan menjadi indikator yang dapat langsung dipakai guru.

Berdasarkan kesenjangan tersebut, kebaruan penelitian ini terletak pada tiga hal. Pertama, kelima prinsip Brown, termasuk kepraktisan yang sering luput dari pembahasan, dipakai sebagai satu kerangka untuk bahasa sekaligus sastra Indonesia. Kedua, disusun matriks yang menghubungkan prinsip, jenis instrumen, dan indikator mutu. Ketiga, keterbatasan setiap instrumen dibahas, tidak hanya keunggulannya. Urgensi penelitian ini terletak pada kebutuhan guru akan acuan yang jelas untuk merancang dan memeriksa mutu instrumen penilaian pada Kurikulum Merdeka, ketika hambatan seperti rubrik yang tidak konsisten dan beban administrasi masih dilaporkan (Wicaksono et al., 2025). Sejalan dengan itu, tujuan penelitian ini adalah (1) mendeskripsikan makna kelima prinsip evaluasi Brown dalam pembelajaran Bahasa dan Sastra Indonesia, (2) menyusun contoh instrumen evaluasi berdasarkan prinsip tersebut, dan (3) mengidentifikasi tantangan penerapannya serta membandingkannya dengan temuan penelitian terdahulu.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan desain studi pustaka (*library research*). Pendekatan kualitatif dipilih karena penelitian ini menelaah makna konsep dan hubungan antarkonsep, bukan menguji hipotesis secara statistik (Creswell & Poth, 2018). Studi pustaka dilakukan sebagai tinjauan literatur semi-sistematis, yaitu menelusuri dan mensintesis berbagai sumber tanpa prosedur seleksi baku seperti pada *systematic review* (Snyder, 2019).

Sumber data terdiri atas data primer dan data sekunder. Data primer berupa satu buku utama, yaitu *Language Assessment: Principles and Classroom Practices* edisi ketiga karya Brown dan (Abeywickrama, 2019). Data sekunder berjumlah 21 sumber, terdiri atas 18 artikel jurnal (11 jurnal internasional dan 7 jurnal nasional), satu artikel prosiding seminar internasional, dan dua buku metodologi penelitian kualitatif. Dengan demikian, jumlah sumber yang ditelaah adalah 22, dan artikel jurnal mencakup 81,8% dari seluruhnya. Seluruh sumber terbit pada rentang 2016–2026.

Sumber dikumpulkan melalui penelusuran daring pada Google Scholar dan laman jurnal ilmiah dengan kata kunci *language assessment principles*, *authentic assessment*, *washback*, *rubric*, *self-assessment*, penilaian autentik, dan evaluasi pembelajaran Bahasa Indonesia. Sumber dipilih dengan tiga kriteria: terbit pada 2016–2026, membahas prinsip atau praktik asesmen (khususnya asesmen bahasa dan pembelajaran Bahasa Indonesia), dan memiliki identitas publikasi yang jelas berupa jurnal, prosiding, atau penerbit. Tulisan berupa opini tanpa dasar kajian tidak dipakai.

Subjek penelitian ini bukan manusia, melainkan konsep, prinsip, dan model evaluasi pembelajaran bahasa. Instrumen utama adalah peneliti sendiri, dibantu lembar analisis literatur untuk mencatat fokus, temuan, dan keterbatasan tiap sumber serta matriks sintesis untuk menghubungkan

prinsip Brown dengan jenis instrumen. Penelitian dilaksanakan sepanjang tahun 2026 dan tidak melibatkan pengumpulan data lapangan. Analisis data mengikuti alur (Miles, et al, 2020) yang disesuaikan untuk kajian pustaka, yaitu kondensasi data, penyajian data, dan penarikan simpulan. Pada tahap kondensasi, setiap sumber dibaca dan bagian yang berkaitan dengan lima prinsip Brown dicatat. Pada tahap penyajian, catatan itu dikelompokkan menurut prinsip dan jenis instrumen ke dalam matriks sintesis (Tabel 1 dan Tabel 2). Pada tahap penarikan simpulan, hasil pengelompokan dibandingkan dengan temuan penelitian terdahulu (Tabel 4). Keabsahan temuan dijaga melalui triangulasi sumber, yaitu membandingkan pendapat dari buku, jurnal internasional, dan jurnal nasional untuk setiap kesimpulan penting (Creswell & Poth, 2018).

HASIL DAN PEMBAHASAN

Hasil Penelitian

Temuan 1: makna lima prinsip Brown dalam pembelajaran Bahasa dan Sastra Indonesia

Telaah terhadap buku utama dan sumber pendukung menunjukkan bahwa kelima prinsip Brown dapat diturunkan menjadi pertanyaan pemeriksa mutu yang konkret bagi guru Bahasa dan Sastra Indonesia (Tabel 1).

Tabel 1.
Penerjemahan Lima Prinsip Evaluasi Brown ke dalam Pembelajaran Bahasa dan Sastra Indonesia

Prinsip	Pengertian menurut Brown	Wujud dalam Bahasa dan Sastra Indonesia	Pertanyaan pemeriksa mutu
Kepraktisan (<i>practicality</i>)	Instrumen efisien dalam waktu dan biaya, mudah dilaksanakan, dan mudah diskor.	Rubrik ringkas untuk presentasi satu kelas; jumlah karya dalam portofolio dibatasi.	Dapatkah guru melaksanakan dan menskor instrumen ini dalam waktu yang tersedia?
Reliabilitas (<i>reliability</i>)	Hasil penilaian konsisten dan dapat diandalkan; sumber ketidakkonsistenan mencakup peserta didik, penilai, pelaksanaan, dan instrumen.	Rubrik analitik dengan deskriptor jelas; karya cerpen dinilai oleh dua penilai.	Apakah dua penilai memberi skor yang berdekatan untuk karya yang sama?
Validitas (<i>validity</i>)	Sejauh mana penafsiran hasil penilaian tepat, bermakna, dan berguna sesuai tujuannya.	Tugas menulis puisi untuk mengukur keterampilan menulis puisi, bukan hafalan unsur intrinsik.	Apakah tugas mengukur kompetensi yang tercantum dalam tujuan pembelajaran?
Autentisitas (<i>authenticity</i>)	Tugas penilaian sepadan dengan tugas penggunaan bahasa di dunia nyata.	Menanggapi cerpen secara lisan, menulis ulasan karya, mementaskan drama.	Apakah peserta didik memakai bahasa untuk tujuan dan pembaca yang nyata?
Dampak balik (<i>washback</i>)	Pengaruh penilaian terhadap pengajaran dan pembelajaran.	Umpan balik atas draf cerpen; lembar refleksi pembelajaran.	Apakah hasil penilaian dipakai peserta didik untuk memperbaiki belajarnya?

Tabel 1 memperlihatkan bahwa setiap prinsip dapat dioperasikan menjadi satu pertanyaan yang bisa dijawab guru sebelum instrumen dipakai. Kepraktisan perlu mendapat tempat yang setara dengan empat prinsip lainnya, sebab instrumen yang baik secara teori tidak akan dipakai jika

membebani guru. Beban administrasi dan keterbatasan waktu termasuk hambatan yang dilaporkan dalam kajian atas penilaian autentik di Indonesia (Wicaksono et al., 2025).

Temuan 2: instrumen evaluasi yang disusun berdasarkan prinsip Brown

Berdasarkan matriks pada Tabel 1, disusun lima contoh instrumen evaluasi. Tabel 2 meringkas hubungan setiap instrumen dengan prinsip yang paling menonjol dan kelemahan yang perlu diantisipasi.

Tabel 2.
Instrumen Evaluasi, Prinsip Brown yang Menonjol, dan Kelemahannya

Instrumen	Bentuk tugas	Prinsip yang menonjol	Kelemahan yang perlu diantisipasi
Presentasi apresiasi sastra	Peserta didik mempresentasikan analisis dan tanggapan atas cerpen atau puisi.	Autentisitas, validitas	Reliabilitas penilai pada penilaian lisan.
Portofolio	Kumpulan cerpen, analisis teks sastra, dan jurnal refleksi.	<i>Washback</i>	Kepraktisan: pemeriksaan memakan waktu.
Proyek pementasan	Kelompok menampilkan drama pendek atau pembacaan puisi.	Autentisitas	Reliabilitas: kontribusi individu sulit dipisahkan.
Menulis kreatif	Tes unjuk kerja menulis cerpen atau puisi.	Validitas	Subjektivitas penilai bila kriteria kabur.
Refleksi diri	Lembar refleksi tertulis.	<i>Washback</i>	Validitas: tidak cukup sebagai satu-satunya bukti kompetensi.

Pada instrumen presentasi apresiasi sastra, peserta didik mempresentasikan hasil analisis dan tanggapannya atas sebuah cerpen atau puisi Indonesia dengan menyinggung tema, amanat, gaya bahasa, dan kaitannya dengan konteks sosial. Tugas ini dekat dengan autentisitas karena peserta didik berbicara kepada pendengar tentang teks yang benar-benar mereka baca, dan dekat dengan validitas karena yang dinilai adalah kemampuan menafsirkan dan menyampaikan, bukan hafalan istilah. Penilaian memakai rubrik analitik dengan empat aspek, yaitu pemahaman teks, ketepatan bahasa, argumentasi, serta sikap dan ekspresi, masing-masing dengan skor 1–4 (Tabel 3).

Tabel 3.
Rubrik Analitik Presentasi Apresiasi Sastra

Aspek	Skor 4	Skor 3	Skor 2	Skor 1
Pemahaman teks	Semua unsur intrinsik yang dibahas tepat dan dikaitkan dengan isi teks.	Sebagian besar unsur intrinsik tepat.	Sebagian kecil unsur tepat; banyak yang keliru.	Unsur intrinsik tidak teridentifikasi atau keliru.
Ketepatan bahasa	Lancar, diksi tepat, kalimat sesuai kaidah.	Cukup lancar; sesekali salah diksi atau struktur.	Sering terhenti; kesalahan mengganggu pemahaman.	Sulit dipahami karena kesalahan bahasa.
Argumentasi	Penafsiran jelas, logis, dan didukung bukti dari teks.	Penafsiran jelas; dukungan bukti terbatas.	Penafsiran kurang jelas; hampir tanpa bukti.	Tidak ada penafsiran yang dapat dinilai.
Sikap dan ekspresi	Percaya diri; intonasi dan sikap mendukung isi.	Cukup percaya diri; intonasi cukup mendukung.	Kurang percaya diri; intonasi datar.	Tidak percaya diri; sulit berkomunikasi.

Mutu rubrik semacam ini ditentukan terutama oleh ketepatan kriterianya dalam menggambarkan capaian belajar yang dituju (Brookhart, 2018). Unsur desain rubrik, seperti kriteria dan deskripsi tiap tingkat kualitas, juga perlu dirumuskan secara eksplisit agar rubrik dapat dipakai dan dilaporkan secara konsisten (Dawson, 2017).

Portofolio memuat kumpulan karya tulis dan refleksi peserta didik dalam satu periode, misalnya sebuah cerpen atau teks narasi, analisis teks sastra, dan jurnal refleksi pembelajaran. Penilaian memperhatikan perkembangan kemampuan menulis, kreativitas dan orisinalitas ide, ketepatan struktur dan kaidah bahasa, serta kedalaman refleksi. Instrumen ini menonjol pada *washback* karena peserta didik dapat melihat perkembangan karyanya dan memperbaikinya berdasarkan umpan balik, sejalan dengan gagasan bahwa penilaian formatif menyatu dengan pedagogi (Black & Wiliam, 2018). Namun, memeriksa portofolio untuk satu kelas penuh memakan waktu, sehingga jumlah karya perlu dibatasi agar kepraktisan terjaga.

Pada penilaian berbasis proyek, peserta didik secara berkelompok menampilkan drama pendek atau pembacaan puisi. Aspek yang dinilai mencakup penghayatan terhadap teks, kerja sama dan tanggung jawab, ketepatan lafal dan intonasi, serta kreativitas penyajian. Tugas ini autentik karena menyerupai kegiatan bersastra di luar kelas. Villarroel et al. (2018) menyebut keterkaitan dengan situasi nyata (*realism*), tantangan kognitif, dan kemampuan peserta didik menilai mutu karya (*evaluative judgement*) sebagai dimensi penilaian autentik. Proyek pementasan memenuhi dimensi pertama dan, bila disertai diskusi tentang kriteria, juga dimensi ketiga. Kelemahannya terletak pada kerja kelompok, sebab nilai kelompok tidak selalu mencerminkan kontribusi tiap anggota sehingga diperlukan catatan observasi individual.

Instrumen menulis kreatif berupa tes unjuk kerja menulis cerpen atau puisi. Indikator penilaiannya adalah kesesuaian tema, struktur teks, gaya bahasa dan imaji, serta orisinalitas gagasan, dengan rubrik analitik berskala 1–4. Validitas menjadi perhatian utama karena tugas harus benar-benar mengukur kemampuan menulis kreatif. Bila tugas terlalu menekankan format, skor lebih mencerminkan kepatuhan pada format daripada kemampuan menulis. Karena karya kreatif bersifat terbuka, penilaian tidak boleh membatasi peserta didik pada jawaban tunggal. Kekhawatiran bahwa rubrik membatasi kreativitas perlu dilihat bersama cara rubrik dirancang dan digunakan (Panadero &

Jonsson, 2020), sehingga kriteria sebaiknya menilai mutu, misalnya keutuhan struktur dan ketepatan imaji, dan bukan kesamaan isi.

Lembar refleksi diri memuat pernyataan seperti “Saya memahami isi dan makna teks sastra yang dipelajari” dan “Saya mampu mengekspresikan pendapat saya secara lisan atau tulisan”, serta kalimat terbuka “Bagian yang paling menantang dalam pembelajaran ini adalah ...”. Meta-analisis menunjukkan bahwa penilaian diri berhubungan positif dengan strategi belajar mandiri dan efikasi diri (Panadero et al., 2017), dan tinjauan lain menyimpulkan bahwa penilaian diri umumnya berkaitan dengan peningkatan belajar, terutama bila peserta didik dibekali kriteria yang jelas (Andrade, 2019). Meski demikian, refleksi diri tidak cukup sebagai satu-satunya bukti kompetensi, sehingga dipakai bersama instrumen lain.

Temuan 3: tantangan penerapan

Telaah sumber menunjukkan empat tantangan. Pertama, konsistensi rubrik antarpemilai. Kajian atas penelitian asesmen autentik di Indonesia melaporkan bahwa rubrik yang tidak seragam antarguru memengaruhi reliabilitas penilaian dan bahwa pelaporan validitas serta reliabilitas instrumen masih jarang (Wicaksono et al., 2025). Kedua, keterbatasan waktu dan beban administrasi (Wicaksono et al., 2025), yang bersinggungan langsung dengan kepraktisan. Ketiga, kapasitas guru. Telaah teoretis tentang penilaian autentik dalam pembelajaran Bahasa Indonesia menekankan perlunya peningkatan kompetensi guru dalam merancang dan melaksanakan penilaian (Friyandini, 2025), dan kajian sistematis tentang evaluasi menulis juga menyebut penguatan kompetensi guru sebagai kebutuhan (Mulyadi et al., 2026). Keempat, sifat interpretatif penilaian sastra, yang menuntut kriteria jelas tetapi tetap memberi ruang bagi perbedaan penafsiran.

Tabel 4 membandingkan fokus dan temuan penelitian ini dengan empat penelitian terdahulu yang paling dekat topiknya.

Tabel 4.
Perbandingan Temuan Penelitian dengan Penelitian Terdahulu

Peneliti	Tahun	Fokus Penelitian	Temuan Utama
Sa'idah et al.	2020	Pengembangan instrumen penilaian otentik berbasis kecakapan hidup pada pembelajaran Bahasa Indonesia	Produk penilaian otentik tergolong baik, reliabel, dan efektif
Wicaksono et al.	2025	Kajian sistematis 16 artikel (2021–2025) tentang asesmen autentik di Indonesia	Asesmen dilakukan lewat proyek, portofolio, refleksi diri, dan penilaian kinerja; hambatan: pelatihan guru, beban administrasi, dukungan sistem
Friyandini	2025	Telaah teoretis asesmen autentik dalam pembelajaran Bahasa Indonesia	Berpotensi meningkatkan keterlibatan dan motivasi; perlu instrumen relevan, kompetensi guru, dan dukungan kebijakan
Mulyadi et al.	2026	Kajian sistematis 10 artikel (2020–2025) tentang evaluasi pembelajaran dan literasi menulis	Evaluasi berperan sentral tetapi masih berfokus pada hasil akhir; perlu penguatan kompetensi guru dan instrumen yang tepat
Penelitian ini	2026	Lima prinsip Brown pada pembelajaran Bahasa dan Sastra Indonesia	Matriks prinsip–instrumen–indikator mutu; tiap instrumen memiliki prinsip menonjol dan kelemahan khas

Penelitian ini memiliki persamaan dengan penelitian terdahulu pada temuan tentang hambatan, yaitu kompetensi guru, rubrik, dan keterbatasan waktu (Friyandini, 2025; Mulyadi et al., 2026; Wicaksono et al., 2025). Persamaan ini justru menguatkan bahwa hambatan tersebut bukan masalah satu sekolah atau satu keterampilan saja. Perbedaannya ada pada tiga hal. Pertama, penelitian terdahulu umumnya memakai asesmen autentik atau literasi menulis sebagai kerangka, sedangkan penelitian ini memakai lima prinsip Brown sehingga validitas, reliabilitas, dan kepraktisan ikut diperiksa. Kedua, penelitian ini mencakup pembelajaran sastra, tidak hanya keterampilan berbahasa. Ketiga, penelitian ini menunjukkan kelemahan khas tiap instrumen, bukan hanya manfaatnya, sehingga guru dapat memperkirakan risiko sebelum memakainya.

Pembahasan

Temuan bahwa setiap instrumen menonjol pada prinsip yang berbeda dapat dijelaskan melalui pandangan bahwa validitas adalah soal kelayakan penafsiran dan penggunaan hasil penilaian, bukan sifat tes semata (Kane, 2016). (Brown & Abeywickrama, 2019) membedakan bukti validitas isi, kriteria, konstruk, konsekuensi, dan tampak (*face validity*). Dalam pembelajaran sastra, bukti isi dapat diperiksa dengan mencocokkan tugas dengan tujuan pembelajaran, bukti konstruk dengan memeriksa apakah rubrik mencerminkan komponen kemampuan yang diukur, dan bukti konsekuensi dengan melihat pengaruh penilaian terhadap cara peserta didik belajar. Cara pandang ini membantu guru menyadari bahwa tes pilihan ganda tentang unsur intrinsik, misalnya, belum tentu mengukur kemampuan mengapresiasi karya.

Pada prinsip reliabilitas, (Brown & Abeywickrama, 2019) membedakan ketidakkonsistenan yang bersumber dari peserta didik, penilai, pelaksanaan, dan instrumen. Pada tugas lisan dan karya kreatif, sumber yang paling menonjol adalah penilai, sehingga rubrik dengan deskriptor yang jelas menjadi kebutuhan, bukan pelengkap. Kajian atas penilaian autentik di Indonesia menyarankan moderasi dan kalibrasi penilaian antarguru melalui diskusi bukti untuk meningkatkan konsistensi (Wicaksono et al., 2025). Saran ini selaras dengan kebutuhan pada instrumen presentasi, menulis kreatif, dan proyek dalam penelitian ini.

Autentisitas dalam kerangka Brown berkaitan dengan kesepadanan tugas penilaian dengan penggunaan bahasa di dunia nyata (Brown & Abeywickrama, 2019). Dimensi realisme, tantangan kognitif, dan penilaian mutu karya oleh peserta didik yang diajukan (Villarroel et al, 2018) memberi cara yang lebih rinci untuk memeriksa autentisitas suatu tugas. Telaah teoretis dalam konteks pembelajaran Bahasa Indonesia melaporkan potensi penilaian autentik dalam meningkatkan keterlibatan dan motivasi (Friyandini, 2025). Meski demikian, autentisitas tinggi tidak otomatis menghasilkan penilaian yang baik. Proyek pementasan, misalnya, sangat autentik tetapi menyulitkan reliabilitas karena kontribusi individu sulit dipisahkan. Di sini terlihat bahwa prinsip-prinsip Brown saling menuntut kompromi.

Prinsip *washback* paling jelas muncul pada portofolio dan refleksi diri. Penilaian yang menyatu dengan pembelajaran memberi informasi yang dapat dipakai guru dan peserta didik untuk memperbaiki proses belajar (Black & Wiliam, 2018). Namun, dampak balik yang baik bergantung pada apakah peserta didik benar-benar memakai umpan balik. (Carless & Boud, 2018) menyebut kemampuan ini sebagai literasi umpan balik, yang mencakup kemampuan menghargai umpan balik, menilai mutu karya, mengelola emosi, dan mengambil tindakan. Karena itu, guru perlu menyediakan waktu bagi peserta didik untuk merevisi karyanya, bukan hanya mengembalikan skor.

Prinsip kepraktisan berkaitan erat dengan kapasitas guru. Kapasitas ini tidak hanya soal keterampilan teknis, tetapi juga cara guru memandang perannya sebagai penilai (Looney et al., 2018). Kajian atas penilaian autentik di Indonesia menyarankan beberapa cara meringankan beban, antara lain portofolio digital dan penilaian berbasis sampel (Wicaksono et al., 2025). Untuk instrumen dalam penelitian ini, jumlah karya portofolio dapat dibatasi dan rubrik yang sama dapat dipakai ulang pada beberapa tugas.

Secara keseluruhan, pembahasan ini menunjukkan bahwa tidak ada satu instrumen yang memenuhi kelima prinsip sekaligus. Kajian sistematis tentang evaluasi menulis juga menyimpulkan bahwa penggunaan satu pendekatan penilaian saja kurang memadai dan perlu dipadukan dengan bentuk penilaian lain (Mulyadi et al., 2026). Dengan demikian, guru sebaiknya memilih beberapa instrumen yang saling menutupi kelemahan, misalnya presentasi atau proyek yang kuat pada autentisitas dipadukan dengan rubrik analitik untuk menjaga reliabilitas dan lembar refleksi untuk memperkuat *washback*.

Penelitian ini memiliki keterbatasan. Kajian bersifat studi pustaka sehingga lima instrumen yang disusun belum diuji di kelas dan belum divalidasi secara empiris, dan jumlah sumber yang ditelaah terbatas pada 22 sumber. Temuan sebaiknya dibaca sebagai kerangka yang perlu diuji lebih lanjut.

KESIMPULAN

Berdasarkan rumusan masalah dan tujuan penelitian, dapat disimpulkan tiga hal. Pertama, kelima prinsip evaluasi Brown, yaitu kepraktisan, reliabilitas, validitas, autentisitas, dan *washback*, dapat dimaknai dalam pembelajaran Bahasa dan Sastra Indonesia sebagai pertanyaan pemeriksa mutu yang konkret, misalnya apakah dua penilai memberi skor yang berdekatan atau apakah hasil penilaian dipakai peserta didik untuk memperbaiki belajarnya. Kedua, kelima prinsip tersebut dapat diwujudkan dalam lima instrumen, yaitu presentasi apresiasi sastra, portofolio, proyek pementasan, menulis kreatif, dan refleksi diri. Setiap instrumen menonjol pada prinsip tertentu dan memiliki kelemahan khas, yaitu reliabilitas penilai pada tugas lisan, beban waktu pada portofolio, kontribusi individu pada proyek kelompok, subjektivitas pada karya kreatif, dan keterbatasan refleksi diri sebagai satu-satunya bukti kompetensi. Karena itu, instrumen perlu dipadukan, bukan dipakai sendiri-sendiri. Ketiga, tantangan penerapan mencakup rubrik yang belum konsisten antarpemilai, keterbatasan waktu dan beban administrasi, kapasitas guru, serta sifat penafsiran karya sastra. Dibandingkan dengan penelitian terdahulu, temuan tentang hambatan bersifat sejalan, sedangkan perbedaannya terletak pada pemakaian lima prinsip Brown sebagai satu kerangka, cakupan pembelajaran sastra, dan penjabaran kelemahan tiap instrumen.

REKOMENDASI

Bagi guru Bahasa dan Sastra Indonesia, disarankan memadukan beberapa instrumen dan menyusun rubrik dengan deskriptor tiap tingkat yang jelas. Bagi sekolah dan pengambil kebijakan, diperlukan pelatihan penilaian yang berkelanjutan serta ruang waktu bagi guru untuk melakukan moderasi penilaian. Bagi peneliti selanjutnya, disarankan menguji kelima instrumen di kelas, menghitung kesepakatan antarpemilai, dan memvalidasinya secara empiris.

DAFTAR PUSTAKA

- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, Article 87. <https://doi.org/10.3389/feduc.2019.00087>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3, Article 22. <https://doi.org/10.3389/feduc.2018.00022>
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360. <https://doi.org/10.1080/02602938.2015.1111294>
- Fadilla, A. R., Suhardi, & Sudiat, S. (2023). Implementasi penilaian autentik Bahasa Indonesia bermuatan literasi digital industri di SMK. *Nuansa Akademik: Jurnal Pembangunan Masyarakat*, 8(2), 277–298.
- Friyandini, R. (2025). Authentic assessment in Indonesian language learning: Theoretical review and its implications. *Proceeding of International Seminar on Humanity, Education, and Language*, 1(1), 397–411. <https://doi.org/10.21009/ishel.v1i1.57721>
- Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198–211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467. <https://doi.org/10.1080/0969594X.2016.1268090>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.
- Mulyadi, U. A., Nadiyah, S., & Hamdani, A. (2026). Evaluasi pembelajaran Bahasa Indonesia terhadap penguatan literasi menulis peserta didik: Tinjauan SLR (systematic literature review). *J-KIP (Jurnal Keguruan dan Ilmu Pendidikan)*, 7(2), 917–928. <https://doi.org/10.25157/j-kip.v7i2.23447>
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, Article 100329. <https://doi.org/10.1016/j.edurev.2020.100329>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>

- Sa'idah, N., Yulistianti, H. D., & Farida, Y. E. (2020). Model penilaian otentik berbasis kecakapan hidup siswa dalam pembelajaran Bahasa Indonesia. *Litera*, 17(2). <https://journal.uny.ac.id/index.php/litera/article/view/18551>
- Sari, R. N. E., & Sukirno, S. (2021). Persepsi guru Bahasa Indonesia SMP Negeri se-Kecamatan Brebes terhadap penilaian autentik untuk keterampilan membaca pada Kurikulum 2013. *Metafora: Jurnal Pembelajaran Bahasa dan Sastra*, 8(1), 1–13.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sunu, E. Y., Cahyono, B. E. H., & Ricahyono, S. (2021). Pelaksanaan penilaian autentik pada mata pelajaran Bahasa Indonesia di SMP Negeri 4 Magetan di masa pandemi Covid-19. *KLITIKA: Jurnal Ilmiah Pendidikan Bahasa dan Sastra Indonesia*, 3(2).
- Susilo, J., Rohmawati, N., & Haryadi, H. (2021). Pengembangan penilaian autentik pada pembelajaran teks eksposisi kelas XI. *Tuturan: Jurnal Pendidikan, Bahasa dan Sastra*, 10(2).
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854. <https://doi.org/10.1080/02602938.2017.1412396>
- Wicaksono, V. D., Rahmawati, I., & Nurlaily, V. A. (2025). Evaluasi sistematis literatur tentang penggunaan asesmen autentik dalam pembelajaran: Studi kasus pendidikan di Indonesia. *Educator: Directory of Elementary Education Journal*, 6(2). <https://doi.org/10.58176/edu.v6i2.3305>

