At-Tadris: Journal of Islamic Education
Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Telaah Pustaka Tentang Manajemen Penilaian Otentik dalam Pembelajaran Bahasa Inggris untuk Mendorong Keterampilan Komunikasi Siswa

Rolando Hanglopo[1], Daflizar[2], Toni Indrayadi[3], Suhaimi[4], Ronal Regen[5]
[1,2,3,4,5]IAIN Kerinci
rolandohanglopo30@gmail.com[1], daflizar@yahoo.com[2], toniindrayadi2@gmail.com[3], suhaimi@iainkerinci.ac.id[4], ronalregern2044@gmail.com[5]

ABSTRACT

The demand for English as a global means of communication positions communicative competence as the central goal of language learning, yet assessment in many classrooms remains oriented toward written tests that measure isolated linguistic knowledge. This article systematically and narratively reviews the conceptual foundations, management framework, implementation forms, and challenges of authentic assessment in English language teaching, and proposes a conceptual model of authentic assessment management that fosters students' communication skills. The study employs a qualitative literature review with thematic analysis of books, journal articles, and relevant regulations. Five findings emerge. First, authentic assessment requires tasks that resemble real-world language use, transparent criteria, and continuous feedback. Second, communication skills are multidimensional, encompassing linguistic, discourse, sociolinguistic, and strategic competence, and therefore cannot be adequately assessed through discrete-item tests. Third, authentic assessment promotes communication through three mechanisms: providing opportunities for interaction, strengthening formative feedback, and generating positive washback. Fourth, the success of authentic assessment depends on management that covers planning, organizing, implementing, and quality control. Fifth, major obstacles such as limited time, class size, teacher assessment literacy, rubric reliability, and learner anxiety can be mitigated through staged task design, inter-rater moderation, and technology integration. The article yields a cyclical model of authentic assessment management and five propositions for further research.

Keywords : *assessment management, authentic assessment, communication skills, English language teaching, literature review.*

ABSTRAK

Tuntutan penguasaan bahasa Inggris sebagai sarana komunikasi global menempatkan keterampilan komunikasi sebagai tujuan utama pembelajaran, sementara praktik penilaian di banyak kelas masih berorientasi pada tes tertulis yang mengukur pengetahuan kebahasaan terisolasi. Artikel ini bertujuan menelaah secara sistematis-naratif landasan konseptual, kerangka manajemen, bentuk implementasi, dan tantangan penilaian otentik dalam pembelajaran bahasa Inggris, serta merumuskan model konseptual manajemen penilaian otentik untuk mendorong keterampilan komunikasi siswa. Metode yang digunakan adalah telaah pustaka kualitatif dengan analisis tematik terhadap buku, artikel jurnal, dan regulasi yang relevan. Hasil telaah menunjukkan lima temuan. Pertama, penilaian otentik menuntut tugas yang menyerupai penggunaan bahasa di dunia nyata, kriteria yang transparan, dan umpan balik yang berkelanjutan. Kedua, keterampilan komunikasi bersifat multidimensi, mencakup kompetensi linguistik, wacana, sosiolinguistik, dan strategis, sehingga tidak dapat dinilai memadai melalui tes butir terpisah. Ketiga, penilaian otentik mendorong komunikasi melalui tiga mekanisme, yaitu pemberian peluang berinteraksi, penguatan umpan balik formatif, dan pengaruh balik (washback) positif terhadap pembelajaran. Keempat, keberhasilan penilaian otentik bergantung pada pengelolaan yang mencakup perencanaan, pengorganisasian, pelaksanaan, dan pengendalian mutu. Kelima, hambatan utama berupa keterbatasan waktu, jumlah siswa, literasi asesmen guru, reliabilitas rubrik, dan kecemasan siswa dapat dimitigasi melalui rancangan tugas bertahap, moderasi

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

antarpenilai, dan pemanfaatan teknologi. Artikel ini menghasilkan model siklus manajemen penilaian otentik dan lima proposisi untuk penelitian lanjutan.

Kata kunci : manajemen penilaian, penilaian otentik, keterampilan komunikasi, pembelajaran bahasa Inggris, telaah Pustaka.

PENDAHULUAN

Bahasa Inggris menempati posisi strategis dalam sistem pendidikan Indonesia karena berfungsi sebagai bahasa pengantar ilmu pengetahuan, teknologi, dan diplomasi, serta sebagai bahasa penghubung lintas negara. Dalam kerangka itu, tujuan akhir pembelajaran bahasa Inggris bukan penguasaan kaidah gramatika semata, melainkan kemampuan menggunakan bahasa untuk berkomunikasi secara lisan dan tertulis dalam beragam situasi sosial. Gagasan ini berakar pada konsep kompetensi komunikatif yang menegaskan bahwa penutur harus menguasai pengetahuan kebahasaan sekaligus kemampuan memakainya secara tepat sesuai konteks, tujuan, dan lawan bicara (Hymes, 1972; Canale & Swain, 1980). Kerangka acuan internasional pun merumuskan kemahiran berbahasa dalam bentuk deskripsi kemampuan melakukan tindakan komunikatif (can-do statements), bukan daftar pengetahuan kebahasaan (Council of Europe, 2020).

Namun demikian, praktik penilaian di sekolah sering kali belum sejalan dengan tujuan komunikatif tersebut. Penilaian yang dominan berupa tes pilihan ganda, isian, atau transformasi kalimat cenderung mengukur pengetahuan gramatikal dan kosakata secara terpisah, sehingga hasilnya kurang menggambarkan kemampuan siswa menggunakan bahasa dalam interaksi nyata. Ketidaksesuaian ini menimbulkan efek pengaruh balik (washback) yang merugikan: guru dan siswa belajar untuk menghadapi tes, bukan untuk berkomunikasi (Messick, 1996; Shepard, 2000). Dalam jangka panjang, siswa mungkin memperoleh nilai baik pada ujian tertulis tetapi gagap ketika diminta berbicara, bernegosiasi, atau menyampaikan gagasan secara spontan.

Penilaian otentik hadir sebagai respons terhadap persoalan tersebut. Wiggins (1990) mengemukakan bahwa penilaian otentik meminta peserta didik menampilkan kinerja yang bermakna dalam konteks yang menyerupai tantangan dunia nyata, bukan sekadar mengingat atau mengenali informasi. Dalam pembelajaran bahasa, penilaian otentik terwujud dalam tugas-tugas seperti simulasi wawancara kerja, bermain peran, presentasi, debat, diskusi kelompok, proyek, pembuatan konten digital, dan portofolio, yang dinilai dengan kriteria eksplisit melalui rubrik. Beberapa penelitian kualitatif di Indonesia melaporkan bahwa guru bahasa Inggris yang diteliti menyambut baik penilaian otentik dan menganggapnya layak diterapkan untuk menilai keterampilan berbicara (Inayah et al., 2019; Al Ayubi et al., 2021).

Meskipun demikian, penerapan penilaian otentik tidak sesederhana mengganti jenis tugas. Penilaian otentik lebih menuntut dibandingkan tes konvensional karena memerlukan perencanaan tugas yang cermat, penyusunan rubrik yang valid, waktu pelaksanaan yang memadai, serta kemampuan guru memberikan umpan balik secara individual. Penelitian terhadap seorang guru bahasa Inggris pada sekolah menengah atas di Bali, misalnya, melaporkan bahwa

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

keterbatasan waktu dan jumlah siswa menyulitkan pemberian kesempatan tampil dan umpan balik bagi setiap siswa, rubrik buatan sendiri belum divalidasi sehingga memunculkan kekhawatiran konsistensi penskoran, dan panduan operasional kurikulum dinilai terbatas (Sukmawati et al., 2026). Kajian sistematis mengenai penilaian berbicara pada konteks bahasa Inggris sebagai bahasa asing juga menyimpulkan bahwa bias penilai, rendahnya literasi asesmen guru, dan kecemasan siswa merupakan tantangan utama, sementara keseimbangan antara keadilan, kepraktisan, dan keotentikan perlu dijaga (Nurkumaila et al., 2025). Persoalan-persoalan ini pada hakikatnya adalah persoalan manajemen: bagaimana penilaian direncanakan, diorganisasikan, dilaksanakan, dan dikendalikan mutunya agar berjalan efektif dan berkelanjutan.

Dari sisi kebijakan, Standar Penilaian Pendidikan menegaskan bahwa penilaian dilakukan oleh pendidik untuk memantau proses, kemajuan, dan hasil belajar, serta untuk memberikan umpan balik bagi peserta didik dan guru (Permendikbudristek Nomor 21 Tahun 2022). Pergeseran kurikulum yang memberi keleluasaan kepada guru merancang asesmen sesuai konteks pembelajaran memperbesar peluang penerapan penilaian otentik, tetapi sekaligus memperbesar tuntutan akan kapasitas manajerial guru dan sekolah.

Berdasarkan penelusuran penulis, sebagian besar kajian terdahulu membahas penilaian otentik dari sudut pandang jenis instrumen atau persepsi guru, sedangkan sudut pandang manajemen, yaitu bagaimana seluruh siklus penilaian dikelola sebagai sistem untuk mendorong keterampilan komunikasi, masih jarang dikaji secara terpadu. Selain itu, kajian yang ada umumnya berskala kasus tunggal sehingga belum memberikan gambaran sintetis lintas konteks. Kesenjangan inilah yang menjadi dasar artikel ini.

Tujuan artikel ini adalah (1) menelaah konsep dan karakteristik penilaian otentik serta hakikat keterampilan komunikasi dalam pembelajaran bahasa Inggris; (2) menganalisis mekanisme yang menghubungkan penilaian otentik dengan peningkatan keterampilan komunikasi; (3) merumuskan kerangka manajemen penilaian otentik yang mencakup perencanaan, pengorganisasian, pelaksanaan, dan pengendalian; serta (4) mengidentifikasi tantangan beserta strategi mitigasinya. Secara teoretis, artikel ini memperkaya kajian asesmen bahasa dengan mengintegrasikan perspektif manajemen. Secara praktis, artikel ini menyediakan rujukan operasional bagi guru, kepala sekolah, dan pengembang kebijakan dalam merancang sistem penilaian yang mendukung kemampuan komunikasi siswa.

METODE PENELITIAN

Penelitian ini menggunakan desain telaah pustaka (literature review) dengan pendekatan kualitatif. Telaah pustaka dipandang tepat apabila tujuan penelitian adalah mensintesis, mengintegrasikan, dan membangun kerangka konseptual dari berbagai sumber yang tersebar (Snyder, 2019). Pendekatan yang dipakai bersifat sistematis-naratif, yakni mengadopsi prinsip transparansi dan keterlacakan dari telaah sistematis (Tranfield et al., 2003) namun menyajikan sintesis secara naratif-tematik karena karakter literatur yang heterogen, mulai dari teori asesmen, studi kasus kualitatif, hingga kajian persepsi guru. Pendekatan kualitatif memungkinkan

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

peneliti menelaah makna, hubungan antarkonsep, dan mekanisme, bukan menghitung besaran efek (Creswell, 2014; Moleong, 2017). Karena itu, hasil artikel ini bersifat sintesis konseptual dan hipotetik, bukan generalisasi statistik.

HASIL DAN PEMBAHASAN

Konseptualisasi Penilaian Otentik

Istilah penilaian otentik (authentic assessment) dipopulerkan Wiggins (1990) sebagai kritik terhadap tes standar yang meminta peserta didik mengenali jawaban dari pilihan yang tersedia. Menurutnya, penilaian dikatakan otentik apabila peserta didik diminta melakukan tugas yang bermakna dan menyerupai tantangan yang dihadapi para praktisi di dunia nyata, dengan kriteria yang diketahui, serta menghasilkan umpan balik yang dapat dipakai untuk memperbaiki kinerja. Wiggins (1998) kemudian menegaskan bahwa penilaian semestinya bersifat edukatif, yakni berfungsi mengembangkan kemampuan dan bukan sekadar menetapkan peringkat. Pandangan ini sejalan dengan ide bahwa asesmen merupakan bagian tak terpisahkan dari pembelajaran, bukan kegiatan yang terpisah di akhir pelajaran (Shepard, 2000).

Gulikers et al. (2004) menawarkan kerangka lima dimensi yang sangat berguna untuk menilai tingkat keotentikan suatu penilaian, yaitu tugas asesmen, konteks fisik dan sosial, bentuk atau hasil kinerja, kriteria dan standar, serta peran peserta didik dalam proses penilaian. Kerangka ini menunjukkan bahwa keotentikan bukanlah label biner, melainkan kontinum. Suatu wawancara kerja simulasi dengan kriteria transparan dan refleksi diri siswa lebih otentik dibandingkan dialog hafalan yang dibacakan di depan kelas, meskipun keduanya sama-sama berupa kegiatan berbicara. Villarroel et al. (2018) menambahkan bahwa keotentikan perlu dipadukan dengan kognisi tingkat tinggi dan kegiatan evaluatif agar penilaian benar-benar menumbuhkan kompetensi, bukan hanya menampilkan realisme permukaan.

Dalam literatur Indonesia, penilaian otentik dipahami sebagai penilaian yang menilai kesiapan, proses, dan hasil belajar secara utuh melalui berbagai teknik, seperti observasi, penilaian kinerja, proyek, portofolio, dan penilaian diri (Kunandar, 2013; Majid, 2014). Penekanan pada penilaian proses dan hasil secara bersamaan penting bagi pembelajaran bahasa karena kemampuan komunikasi tumbuh bertahap dan tampak dalam rangkaian kinerja, bukan dalam satu kali tes.

Berdasarkan telaah, penilaian otentik dalam pembelajaran bahasa memiliki lima karakteristik inti. Pertama, tugas bersifat kontekstual dan bertujuan komunikatif, misalnya menyampaikan pendapat, meminta informasi, atau menegosiasikan solusi. Kedua, tugas menuntut integrasi keterampilan, sehingga berbicara dipadukan dengan menyimak, membaca, dan menulis. Ketiga, kinerja dinilai dengan kriteria eksplisit yang diketahui siswa sejak awal. Keempat, penilaian menyediakan umpan balik yang segera dan dapat ditindaklanjuti. Kelima, siswa dilibatkan sebagai pelaku aktif melalui penilaian diri dan penilaian sejawat. Tabel 1 merangkum perbedaan penilaian otentik dengan tes konvensional.

At-Tadris: Journal of Islamic Education
Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Tabel 1. Perbandingan tes konvensional dan penilaian otentik dalam pembelajaran bahasa Inggris

Aspek	Tes konvensional	Penilaian otentik
Fokus pengukuran	Pengetahuan kebahasaan terpisah (kosakata, gramatika)	Penggunaan bahasa dalam tugas komunikatif bermakna
Bentuk tugas	Pilihan ganda, isian, transformasi kalimat	Bermain peran, wawancara, presentasi, diskusi, proyek, portofolio
Konteks	Dekontekstual dan seragam	Kontekstual, menyerupai situasi nyata
Kriteria	Kunci jawaban benar-salah	Rubrik analitik atau holistik yang transparan
Umpan balik	Umumnya berupa skor akhir	Berkelanjutan, deskriptif, dan dapat ditindaklanjuti
Peran siswa	Penjawab pasif	Pelaku aktif, termasuk penilaian diri dan sejawat
Pengaruh balik	Cenderung mengarahkan belajar untuk tes	Cenderung mengarahkan belajar untuk berkomunikasi

Meskipun penilaian otentik memiliki kelebihan, kajian yang berimbang juga perlu menyebut tuntutan psikometrisnya. Pada tes konvensional, reliabilitas relatif mudah dicapai melalui penskoran objektif. Pada penilaian kinerja, penskoran melibatkan pertimbangan penilai sehingga reliabilitas dan keadilan harus dikelola secara sengaja, antara lain melalui rubrik yang teruji dan pelatihan penilai (Jonsson & Svingby, 2007; Fulcher, 2010). Penilaian otentik yang baik karenanya tidak menolak ukuran mutu, melainkan menyeimbangkan validitas konstruk dan keotentikan dengan reliabilitas dan kepraktisan, sebagaimana dirumuskan dalam konsep kebermanfaatan tes (test usefulness) oleh Bachman dan Palmer (1996).

Hakikat Keterampilan Komunikasi dalam Bahasa Inggris

Keterampilan komunikasi dalam pembelajaran bahasa Inggris bertumpu pada konsep kompetensi komunikatif. Hymes (1972) mengkritik pandangan yang hanya menekankan kompetensi gramatikal dan menegaskan perlunya kemampuan mengetahui kapan, di mana, dan bagaimana bahasa dipakai secara layak. Canale dan Swain (1980) menjabarkan kompetensi komunikatif ke dalam kompetensi gramatikal, sosiolinguistik, dan strategis, dan Canale (1983) menambahkan kompetensi wacana. Celce-Murcia et al. (1995) selanjutnya mengembangkan model yang memuat kompetensi wacana sebagai pusat dan kompetensi linguistik, aksional, sosiokultural, serta strategis sebagai penopangnya.

Bachman dan Palmer (2010) memandang kemampuan berbahasa sebagai kombinasi pengetahuan bahasa dan kompetensi strategis yang diaktifkan dalam penggunaan bahasa untuk tujuan tertentu. Acuan Eropa untuk bahasa (Council of

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Europe, 2020) menekankan bahwa kemampuan komunikatif tampak dalam empat mode, yaitu reseptif, produktif, interaktif, dan mediasi. Penekanan pada mode interaktif dan mediasi relevan bagi penilaian di kelas, karena komunikasi sehari-hari lebih banyak berupa percakapan dua arah dan pengalihan makna daripada monolog yang disiapkan.

Dengan memadukan ketiga pandangan tersebut, keterampilan komunikasi siswa dalam artikel ini dioperasionalkan menjadi lima dimensi yang dapat diamati. Dimensi pertama adalah kelancaran (fluency), yaitu kemampuan menyampaikan gagasan secara mengalir tanpa jeda yang mengganggu. Dimensi kedua adalah ketepatan kebahasaan (accuracy), meliputi gramatika, kosakata, dan pelafalan yang mendukung keterpahaman. Dimensi ketiga adalah kemampuan interaksi, yakni menanggapi lawan bicara, mengambil giliran, dan menegosiasikan makna. Dimensi keempat adalah kompetensi sosiolinguistik dan wacana, yaitu ketepatan register, kesantunan, serta kohesi dan koherensi gagasan. Dimensi kelima adalah kompetensi strategis, termasuk kemampuan memperbaiki kesalahpahaman, memparafrasekan, dan mengatasi kekurangan kosakata. Kelima dimensi ini menunjukkan bahwa penilaian komunikasi tidak cukup dengan menghitung kesalahan gramatika, tetapi memerlukan tugas yang memaksa siswa menggunakan seluruh dimensi tersebut sekaligus (Luoma, 2004).

Satu catatan penting menyangkut sisi afektif. Kecemasan berbahasa asing merupakan faktor yang secara konsisten memengaruhi kemauan berkomunikasi dan kinerja lisan (Horwitz et al., 1986). Penilaian komunikasi yang dirancang tanpa mempertimbangkan kecemasan dapat menghasilkan skor yang menggambarkan keadaan emosional siswa lebih daripada kemampuan bahasanya. Oleh karena itu dimensi afektif, yang mencakup kepercayaan diri dan kemauan berkomunikasi, perlu dipertimbangkan dalam rancangan tugas dan iklim penilaian, meskipun tidak selalu dijadikan komponen skor.

Keterkaitan antara dimensi komunikasi dan jenis tugas perlu dipetakan secara sadar. Dimensi kelancaran paling mudah tampak pada presentasi dan penceritaan, karena siswa memiliki ruang untuk mengembangkan ujaran yang panjang. Dimensi interaksi hanya dapat diamati pada tugas yang menuntut tanggapan spontan, seperti diskusi, wawancara, dan negosiasi, sehingga presentasi hafalan tidak memadai untuk menilainya. Dimensi sosiolinguistik muncul ketika tugas menetapkan peran dan hubungan sosial yang berbeda, misalnya berbicara kepada guru dibandingkan kepada teman. Dimensi strategis terlihat ketika tugas memuat kesenjangan informasi atau tantangan yang memaksa siswa memperbaiki kesalahpahaman. Pemetaan ini membantu guru memilih rangkaian tugas yang secara kolektif menutup seluruh dimensi, sehingga skor akhir tidak bias pada satu aspek saja. Prinsip serupa dikenal dalam pengembangan asesmen bahasa, yaitu bahwa spesifikasi tugas harus diturunkan dari konstruk yang hendak diukur (Bachman & Palmer, 2010; Luoma, 2004). Dengan kata lain, keotentikan tugas dan kecermatan konstruk harus berjalan bersama.

At-Tadris: Journal of Islamic Education
Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Mekanisme Penilaian Otentik dalam Mendorong Keterampilan Komunikasi

Telaah teoretis menunjukkan sedikitnya tiga mekanisme yang menghubungkan penilaian otentik dengan peningkatan keterampilan komunikasi.

Mekanisme pertama adalah penyediaan peluang berinteraksi dan menghasilkan bahasa. Teori interaksi menegaskan bahwa negosiasi makna dalam percakapan mendorong pemerolehan bahasa (Long, 1996), sementara hipotesis keluaran komprehensif (comprehensible output) menyatakan bahwa upaya memproduksi bahasa memaksa pembelajar memperhatikan bentuk dan memperbaiki ujarannya (Swain, 1985). Penilaian otentik berbasis tugas, seperti diskusi berpasangan atau bermain peran, secara struktural menyediakan situasi tersebut. Pandangan sosiokultural juga menempatkan interaksi sosial sebagai sumber perkembangan kemampuan, dengan bantuan lawan bicara atau guru pada zona perkembangan proksimal (Vygotsky, 1978). Penilaian yang menempatkan siswa dalam kerja pasangan terbukti memberi rasa nyaman bagi siswa pemalu untuk berbicara, meskipun penelitian yang sama mencatat bahwa pasangan dapat didominasi oleh siswa yang lebih percaya diri sehingga partisipasi tidak selalu setara (Sukmawati et al., 2026). Temuan ini menunjukkan bahwa peluang berinteraksi memerlukan pengelolaan agar adil.

Mekanisme kedua adalah penguatan umpan balik formatif. Umpan balik merupakan salah satu faktor yang berdampak besar terhadap pembelajaran apabila menjelaskan capaian, kesenjangan, dan langkah perbaikan (Hattie & Timperley, 2007). Black dan Wiliam (1998) menunjukkan bahwa asesmen formatif yang menghasilkan umpan balik bermakna meningkatkan capaian belajar, terutama bagi siswa yang berkemampuan rendah. Sadler (1989) menekankan bahwa siswa memerlukan pemahaman tentang standar kualitas dan kemampuan membandingkan karyanya dengan standar tersebut. Penilaian otentik memenuhi kondisi ini karena kriterianya transparan dan kinerjanya dapat diamati langsung, sehingga umpan balik dapat bersifat spesifik, misalnya pada pilihan kata untuk register formal atau strategi memperbaiki salah paham. Pada penelitian tentang strategi penilaian berbicara yang telah disebutkan, guru menilai kelancaran, pelafalan, kosakata, tata bahasa, dan kepercayaan diri dengan rubrik analitik serta memberikan umpan balik langsung (Sukmawati et al., 2026), yang menunjukkan bentuk operasional mekanisme ini.

Mekanisme ketiga adalah pengaruh balik positif terhadap pembelajaran (positive washback). Messick (1996) menyatakan bahwa tes memengaruhi cara guru mengajar dan siswa belajar, dan pengaruh itu menjadi bukti validitas konsekuensial. Apabila tugas penilaian berupa komunikasi nyata, guru terdorong memperbanyak latihan komunikasi di kelas dan siswa terdorong mempraktikkannya. Sebaliknya, apabila penilaian berupa soal pilihan ganda, pembelajaran cenderung menyempit pada latihan soal. Dengan demikian penilaian otentik tidak hanya mengukur keterampilan komunikasi, tetapi juga membentuknya melalui arah pembelajaran yang ditimbulkannya. Kajian tentang praktik penilaian berbicara melaporkan bahwa pendekatan berbasis kinerja dan berbantuan teknologi merupakan pendekatan yang paling umum, dan bahwa penilaian formatif yang didukung teknologi dapat memperbaiki umpan balik dan keterlibatan siswa (Nurkumaila et al., 2025).

At-Tadris: Journal of Islamic Education
Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Ketiga mekanisme tersebut tidak berjalan otomatis. Mekanisme interaksi tergantung pada rancangan tugas, mekanisme umpan balik tergantung pada kapasitas dan waktu guru, dan mekanisme washback tergantung pada selarasnya penilaian dengan tujuan pembelajaran. Ketergantungan inilah yang menjadi alasan konseptual bahwa penilaian otentik perlu dikelola secara sistematis, yang diuraikan pada subbab berikut.

Kerangka Manajemen Penilaian Otentik

Manajemen pada dasarnya adalah proses mengoordinasikan sumber daya untuk mencapai tujuan melalui fungsi perencanaan, pengorganisasian, pelaksanaan, dan pengendalian (Terry & Rue, 2010). Penerapan fungsi tersebut pada penilaian otentik menghasilkan kerangka yang dirangkum pada Tabel 2 dan diuraikan berikut.

Tabel 2. Kerangka manajemen penilaian otentik untuk keterampilan komunikasi

Fungsi	Kegiatan kunci	Keluaran
Perencanaan	Merumuskan tujuan komunikatif, memetakan capaian pembelajaran, merancang tugas dan rubrik, menjadwalkan penilaian	Peta asesmen, deskripsi tugas, rubrik
Pengorganisasian	Membagi peran guru dan siswa, mengatur ruang, waktu, kelompok, dan sarana, membangun kolaborasi antarguru	Jadwal, pengelompokan, tim penilai
Pelaksanaan	Menyajikan tugas, mengamati dan merekam kinerja, memberi umpan balik formatif, melibatkan penilaian diri dan sejawat	Catatan kinerja, rekaman, portofolio
Pengendalian	Moderasi antarpenilai, telaah validitas rubrik, analisis hasil, tindak lanjut pembelajaran	Skor yang dapat dipertanggungjawabkan, rencana perbaikan

Perencanaan

Perencanaan merupakan fungsi yang paling menentukan karena kualitas penilaian otentik lebih banyak ditentukan oleh rancangan tugas daripada oleh pelaksanaan di kelas. Langkah pertama adalah merumuskan tujuan komunikatif yang spesifik, misalnya menyampaikan pendapat dan alasannya, menegosiasikan rencana kegiatan, atau memberikan instruksi. Rumusan ini diturunkan dari capaian pembelajaran dan diselaraskan dengan standar penilaian yang menuntut kesesuaian antara tujuan, teknik, dan instrumen (Permendikbudristek Nomor 21 Tahun 2022). Prinsip keselarasan konstruktif menegaskan bahwa tujuan, kegiatan belajar, dan penilaian harus saling mendukung; tugas yang tidak selaras akan menghasilkan skor yang tidak valid (Biggs, 1996).

Langkah kedua adalah merancang tugas yang otentik tetapi layak. Dengan menggunakan lima dimensi Gulikers et al. (2004), guru dapat memeriksa apakah

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

tugas memiliki situasi yang menyerupai kehidupan nyata, peran sosial yang jelas, produk kinerja yang dapat diamati, kriteria yang diketahui, dan ruang bagi siswa untuk merefleksikan hasil. Pembelajaran berbasis tugas (task-based language teaching) menyediakan prinsip perancangan yang relevan, yaitu tugas memiliki tujuan nonkebahasaan, kesenjangan informasi atau pendapat, dan hasil yang dapat dinilai (Ellis, 2003; Nunan, 2004). Langkah ketiga adalah menyusun rubrik. Rubrik analitik membantu menjelaskan harapan, meningkatkan konsistensi penskoran, dan mendukung umpan balik, meskipun manfaatnya bergantung pada kualitas rancangan dan cara penggunaannya (Brookhart, 2013; Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). Rubrik yang baik memuat deskriptor perilaku yang dapat diamati pada setiap tingkat, bukan istilah evaluatif yang samar seperti baik atau kurang. Langkah keempat adalah penjadwalan: tugas dipecah menjadi beberapa tahap kecil sepanjang semester sehingga tidak menumpuk pada satu periode ujian.

Pengorganisasian

Pengorganisasian berkaitan dengan penataan sumber daya manusia, waktu, ruang, dan sarana. Persoalan paling nyata adalah keterbatasan waktu dan jumlah siswa. Penelitian pada satu kelas sekolah menengah atas melaporkan bahwa waktu pelajaran yang terbatas membuat tidak semua siswa mendapat kesempatan tampil dan umpan balik individual (Sukmawati et al., 2026). Strategi pengorganisasian yang dapat ditempuh antara lain pembagian kelas menjadi stasiun, pelaksanaan penilaian secara bergilir dalam beberapa pertemuan, pemanfaatan kerja berpasangan dan berkelompok, serta pengumpulan kinerja berupa rekaman audio atau video di luar jam pelajaran.

Pengorganisasian juga mencakup kolaborasi. Guru dapat berbagi rubrik, melakukan penskoran silang, dan menyelenggarakan pertemuan moderasi melalui forum guru mata pelajaran. Kepala sekolah berperan menyediakan waktu dan insentif bagi kolaborasi tersebut, serta memastikan pelatihan literasi asesmen. Literasi asesmen, yakni pengetahuan dan keterampilan guru dalam merancang, menggunakan, dan menafsirkan asesmen (Stiggins, 1991; Popham, 2009), merupakan sumber daya utama. Penelitian yang dikaji menemukan bahwa guru yang tidak memperoleh pelatihan formal penilaian menyusun rubrik tanpa validasi sehingga muncul kekhawatiran tentang konsistensi dan keandalan skor (Sukmawati et al., 2026), dan kajian sistematis merekomendasikan penguatan literasi asesmen guru agar mampu menyusun rubrik, memberikan umpan balik formatif, dan menafsirkan skor berbicara dengan tepat (Nurkumaila et al., 2025).

Pelaksanaan

Pelaksanaan penilaian otentik sebaiknya mengikuti urutan yang memberi dukungan bertahap. Tahap persiapan memperkenalkan tugas, tujuan, dan rubrik kepada siswa, memberikan contoh kinerja pada berbagai tingkat, serta menyediakan latihan. Tahap kinerja memberikan situasi komunikatif seperti wawancara, bermain peran, atau diskusi, sementara guru mengamati dan mencatat. Tahap umpan balik menyampaikan hasil secara deskriptif dan memberi kesempatan memperbaiki. Pada tahap akhir, siswa melakukan refleksi dan menetapkan sasaran perbaikan.

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Jenis tugas dapat disesuaikan dengan tujuan komunikatif. Wawancara dan simulasi cocok untuk komunikasi interaktif, presentasi untuk komunikasi produktif monolog, diskusi dan debat untuk negosiasi makna dan argumentasi, proyek dan pembuatan konten digital untuk integrasi keterampilan, serta portofolio untuk menunjukkan perkembangan dari waktu ke waktu. Penelitian di Indonesia melaporkan penggunaan tugas wawancara berpasangan, presentasi individu, dan bermain peran dengan situasi kehidupan nyata, seperti memesan makanan di restoran, dengan penilaian atas kelancaran, pelafalan, kosakata, tata bahasa, dan kepercayaan diri (Sukmawati et al., 2026). Studi deskriptif kualitatif pada seorang guru dan 28 siswa sekolah menengah pertama di Banda Aceh melaporkan bahwa guru menilai sikap melalui observasi, pengetahuan melalui wawancara atau pertanyaan singkat, dan keterampilan melalui kegiatan menarasikan urutan peristiwa, dan para penulisnya menyimpulkan bahwa penilaian otentik merupakan cara yang layak untuk menilai berbicara (Inayah et al., 2019). Adapun studi kasus tentang persepsi dua guru menyoroti kesulitan mengelola kegiatan kelas, menjaga konsistensi penilaian, dan memastikan keaslian pekerjaan siswa (Al Ayubi et al., 2021).

Pengendalian Mutu

Pengendalian mutu memastikan bahwa skor dapat dipertanggungjawabkan dan hasil penilaian ditindaklanjuti. Moderasi antarpenilai, yaitu kegiatan beberapa penilai menskor sampel kinerja yang sama lalu membandingkan serta menyepakati tafsir deskriptor, merupakan strategi sederhana untuk mengurangi bias penilai yang disebut sebagai tantangan utama dalam penilaian berbicara (Nurkumaila et al., 2025). Validitas rubrik dapat ditelaah melalui penelaahan sejawat dan uji coba terbatas. Analisis hasil perlu menghasilkan tindak lanjut berupa pembelajaran perbaikan, pengayaan, atau penyesuaian tugas. Selain itu, guru perlu meninjau validitas konsekuensial, yaitu apakah penilaian benar-benar mengarahkan pembelajaran pada komunikasi (Messick, 1996). Siklus tertutup ini yang membedakan manajemen penilaian dari sekadar pelaksanaan tugas.

Integrasi Teknologi dalam Siklus Manajemen

Teknologi dapat berperan pada setiap fungsi manajemen. Pada tahap perencanaan, guru dapat memanfaatkan perangkat daring untuk menyusun dan membagikan rubrik serta contoh kinerja. Pada tahap pengorganisasian, penjadwalan dan pengumpulan tugas dapat dilakukan melalui platform pembelajaran. Pada tahap pelaksanaan, perekaman audio dan video memungkinkan siswa menyampaikan kinerja di luar jam pelajaran sehingga waktu tatap muka digunakan untuk interaksi dan umpan balik. Pada tahap pengendalian, rekaman memudahkan penskoran ulang dan moderasi karena penilai dapat meninjau kinerja yang sama secara berulang. Kajian sistematis menyimpulkan bahwa asesmen berbasis kinerja dan berbantuan teknologi merupakan pendekatan yang paling lazim dalam penilaian berbicara dan berpotensi memperbaiki umpan balik serta keterlibatan siswa (Nurkumaila et al., 2025).

At-Tadris: Journal of Islamic Education
Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Meskipun demikian, teknologi bukan jaminan mutu. Rekaman tidak menggantikan kriteria yang baik, dan kemudahan perangkat tidak otomatis meningkatkan keotentikan tugas. Guru perlu memastikan bahwa teknologi dipilih karena tujuan komunikatif, bukan karena kebaruan alat. Aspek keadilan juga menuntut perhatian, sebab ketimpangan akses perangkat dan konektivitas dapat menyebabkan siswa tertentu dirugikan. Kajian pada seorang guru sekolah menengah atas menunjukkan bahwa teknologi tersedia tetapi jarang digunakan karena keterbatasan keterampilan digital guru (Sukmawati et al., 2026), yang menegaskan bahwa kesiapan manusia sama pentingnya dengan ketersediaan alat. Selain itu, kemunculan perangkat kecerdasan buatan generatif menuntut guru mendesain tugas yang menekankan kinerja langsung, interaksi spontan, dan refleksi proses sehingga keaslian kemampuan siswa tetap dapat dinilai (Al Ayubi et al., 2021).

Rancangan Rubrik Penilaian Komunikasi

Rubrik merupakan instrumen sentral karena menerjemahkan keotentikan tugas menjadi kriteria yang dapat diamati. Tabel 3 menyajikan contoh dimensi rubrik analitik yang dapat diadaptasi untuk penilaian berbicara interaktif, disusun dengan mengacu pada dimensi kompetensi komunikatif pada subbab 3.2. Skala dapat dikalibrasi sesuai jenjang dan capaian pembelajaran.

Tabel 3. Contoh dimensi dan deskriptor rubrik analitik penilaian berbicara interaktif

Dimensi	Deskriptor tingkat tinggi	Deskriptor tingkat berkembang
Kelancaran	Menyampaikan gagasan mengalir dengan jeda yang tidak mengganggu	Sering terhenti sehingga alur gagasan terputus
Ketepatan kebahasaan	Kesalahan jarang dan tidak mengganggu keterpahaman	Kesalahan sering dan kadang mengaburkan makna
Interaksi	Menanggapi, mengambil giliran, dan menegosiasikan makna secara aktif	Bergantung pada lawan bicara dan jarang merespons
Sosiolinguistik dan wacana	Register dan kesantunan sesuai konteks, gagasan runtut	Register kurang tepat, gagasan kurang terhubung
Strategi komunikasi	Memperbaiki salah paham dengan parafrasa atau klarifikasi	Berhenti ketika kekurangan kosakata

Beberapa prinsip perlu diperhatikan dalam penggunaan rubrik. Pertama, jumlah dimensi sebaiknya terbatas agar penilai dapat menggunakannya secara konsisten saat mengamati kinerja langsung. Kedua, deskriptor dirumuskan secara positif dan berbasis perilaku sehingga berfungsi sebagai pedoman perbaikan, bukan sekadar pembenaran skor. Ketiga, rubrik dibagikan kepada siswa sebelum tugas dan dipakai untuk penilaian diri dan sejawat, sebab pemahaman terhadap kriteria mendorong kemampuan regulasi diri (Panadero & Jonsson, 2013). Keempat, rubrik

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

diuji coba dan direvisi berdasarkan hasil moderasi. Rubrik yang tidak divalidasi, seperti yang ditemukan pada praktik seorang guru yang diteliti, berisiko menimbulkan inkonsistensi penskoran (Sukmawati et al., 2026).

Penilaian Diri, Penilaian Sejawat, dan Portofolio

Tiga bentuk penilaian partisipatif berperan penting dalam menjaga keberlanjutan penilaian otentik pada kelas yang besar. Penilaian diri mendorong siswa membandingkan kinerjanya dengan kriteria dan menetapkan sasaran perbaikan, sehingga memperkuat regulasi diri dalam belajar (Zimmerman, 2002; Boud & Falchikov, 2007). Penilaian sejawat memperluas sumber umpan balik dan melatih siswa menerapkan kriteria; namun kualitasnya bergantung pada pelatihan dan pengawasan guru agar tidak bias oleh pertemanan atau ketidakpercayaan diri (Topping, 1998). Portofolio mengumpulkan sampel kinerja dari waktu ke waktu beserta refleksi siswa, sehingga menggambarkan perkembangan kemampuan komunikasi yang tidak tertangkap oleh satu kali penilaian (Paulson et al., 1991).

Dalam kerangka manajemen, ketiga bentuk ini dapat diposisikan sebagai pengungkit efisiensi sekaligus pedagogis. Efisiensi tercapai karena sebagian beban umpan balik dibagi kepada siswa, sedangkan manfaat pedagogisnya adalah siswa menginternalisasi standar kualitas (Sadler, 1989). Guru tetap memegang tanggung jawab akhir atas penilaian sumatif, sedangkan penilaian diri dan sejawat lebih tepat dipakai untuk fungsi formatif. Prinsip ini menjaga keseimbangan antara partisipasi siswa dan akuntabilitas penilaian.

Ilustrasi Penerapan Siklus Manajemen pada Satu Unit Pembelajaran

Agar kerangka manajemen lebih operasional, berikut ilustrasi konseptual pada satu unit tentang menyampaikan pendapat dan memberi saran pada tingkat sekolah menengah. Ilustrasi ini disusun penulis berdasarkan sintesis literatur dan bukan merupakan data empiris.

Pada tahap perencanaan, guru menetapkan tujuan bahwa siswa mampu mendiskusikan suatu masalah di lingkungan sekolah dan menyepakati saran perbaikan bersama pasangan. Guru merancang tugas berupa diskusi rapat kecil dengan peran yang jelas, menyusun rubrik berlima dimensi seperti pada Tabel 3, dan menjadwalkan latihan pada pertemuan pertama, penilaian formatif pada pertemuan kedua, serta penilaian bernilai pada pertemuan ketiga. Pada tahap pengorganisasian, kelas dibagi ke dalam beberapa kelompok yang tampil bergiliran, sementara kelompok lain bertugas sebagai pengamat yang mengisi lembar penilaian sejawat. Guru berkoordinasi dengan rekan sejawat untuk saling meninjau rubrik dan menyepakati tafsir deskriptor.

Pada tahap pelaksanaan, guru memperkenalkan rubrik dan menunjukkan contoh kinerja, kemudian siswa berlatih. Pada penilaian formatif, guru mencatat kinerja dan memberikan umpan balik spesifik, misalnya pada penggunaan ungkapan memberi saran yang sopan atau strategi meminta klarifikasi. Pada penilaian bernilai, kinerja direkam bila memungkinkan. Siswa menutup unit dengan refleksi tentang dimensi yang telah berkembang dan sasaran berikutnya. Pada tahap pengendalian, dua guru menskor sampel rekaman yang sama untuk melihat kesepakatan,

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

membahas perbedaan tafsir, lalu merevisi deskriptor yang ambigu. Hasil dianalisis untuk mengidentifikasi dimensi yang masih lemah, misalnya interaksi, dan guru merancang kegiatan penguatan pada unit berikutnya. Ilustrasi ini memperlihatkan bahwa penilaian otentik yang dikelola secara siklis menghasilkan data yang bermanfaat bagi pembelajaran, bukan sekadar skor.

Keadilan, Etika, dan Inklusivitas Penilaian

Penilaian otentik menimbulkan pertimbangan keadilan yang khas. Karena kinerja dinilai melalui pertimbangan penilai dan sering berlangsung di depan publik, risiko bias lebih besar dibandingkan tes objektif. Standar pengujian menekankan bahwa keadilan menuntut kesempatan yang setara bagi seluruh peserta didik untuk menunjukkan kemampuannya dan interpretasi skor yang sahih bagi semua kelompok (American Educational Research Association et al., 2014). Dalam praktik kelas, hal ini berarti guru perlu menyediakan informasi dan latihan yang setara, mempertimbangkan latar belakang bahasa dan akses sumber daya, serta memberi alternatif bagi siswa yang mengalami hambatan, misalnya kecemasan berat atau kebutuhan khusus.

Aspek etika juga mencakup penggunaan hasil penilaian. Tes memiliki kekuatan sosial yang dapat mengubah perilaku dan peluang seseorang (Shohamy, 2001), sehingga penilaian yang berdampak tinggi memerlukan bukti validitas yang lebih kuat. Pada penilaian kelas yang bersifat formatif, risiko tersebut lebih rendah, tetapi guru tetap perlu menghindari penggunaan hasil kinerja lisan untuk mempermalukan siswa. Pembingkaian kesalahan sebagai bagian wajar dari proses belajar penting untuk menjaga iklim yang aman dan mendorong keberanian berkomunikasi. Dalam kerangka validitas berbasis argumen, setiap keputusan penilaian perlu dapat dibela dengan bukti bahwa skor menggambarkan kemampuan komunikasi yang dimaksud dan bahwa konsekuensinya dapat diterima (Kane, 2006; Bachman & Palmer, 2010).

Peran Kepemimpinan dan Budaya Asesmen Sekolah

Penilaian otentik tidak dapat berkelanjutan apabila hanya bertumpu pada inisiatif individu guru. Kepemimpinan sekolah berpengaruh pada mutu pengajaran melalui penciptaan prioritas, dukungan sumber daya, dan budaya profesional (Hallinger & Heck, 1998). Dalam konteks asesmen, kepala sekolah dapat memastikan tersedianya waktu moderasi, menyetujui pembagian beban penilaian yang realistis, memfasilitasi pelatihan, serta mengapresiasi praktik baik. Budaya asesmen yang sehat ditandai oleh keterbukaan guru berbagi rubrik dan sampel kinerja, kebiasaan berdiskusi tentang bukti belajar siswa, dan pandangan bahwa asesmen adalah sarana perbaikan, bukan sarana penghakiman. Temuan bahwa panduan operasional kurikulum yang terbatas menimbulkan variasi praktik antarguru (Sukmawati et al., 2026) menunjukkan bahwa dukungan struktural pada tingkat sekolah dan dinas pendidikan penting untuk menjaga kualitas bersama.

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Perluasan pada Keterampilan Bahasa Lain

Meskipun artikel ini berfokus pada komunikasi lisan, prinsip manajemen penilaian otentik dapat diperluas pada keterampilan bahasa lainnya karena komunikasi nyata hampir selalu melibatkan beberapa keterampilan sekaligus. Pada keterampilan menyimak, tugas otentik dapat berupa mendengarkan pengumuman, penjelasan, atau percakapan lalu menindaklanjutinya, misalnya mencatat informasi penting dan merespons dengan tindakan atau pertanyaan klarifikasi. Pada keterampilan membaca, tugas dapat berupa membaca beberapa sumber untuk memutuskan pilihan, mengisi kesenjangan informasi, atau merangkum bagi pembaca tertentu. Pada keterampilan menulis, tugas dapat berupa surel formal, proposal kegiatan, ulasan, atau poster informasi yang disusun melalui tahap rancangan, umpan balik, dan revisi (Brown & Abeywickrama, 2019).

Tugas terintegrasi, misalnya membaca dua teks tentang suatu isu, mendiskusikannya dengan pasangan, lalu menulis rekomendasi, mencerminkan cara bahasa dipakai di dunia nyata dan sejalan dengan penekanan acuan Eropa pada mode interaktif dan mediasi (Council of Europe, 2020). Dari sisi manajemen, tugas terintegrasi lebih efisien karena satu rangkaian tugas dapat menghasilkan bukti bagi beberapa capaian pembelajaran, sehingga mengurangi jumlah kegiatan penilaian terpisah. Konsekuensinya, rubrik perlu dirancang agar dapat memisahkan kontribusi setiap keterampilan, atau secara eksplisit menyatakan bahwa yang dinilai adalah kinerja gabungan. Tanpa kejelasan ini, skor akhir akan sulit ditafsirkan dan umpan balik kehilangan ketepatannya.

Sintesis Temuan Empiris Terdahulu di Indonesia

Untuk memetakan landasan empiris, Tabel 4 merangkum temuan penelitian terdahulu yang relevan, yang sebagian besar berfokus pada penilaian berbicara dalam bahasa Inggris. Perlu ditegaskan bahwa sebagian ringkasan didasarkan pada abstrak dan bagian artikel yang dapat diakses penulis, sehingga uraian dibatasi pada hal yang dilaporkan secara eksplisit.

Tabel 4. Ringkasan temuan penelitian terdahulu tentang penilaian otentik dalam pembelajaran bahasa Inggris

Penulis dan tahun	Konteks dan fokus	Temuan utama
Inayah et al. (2019)	Satu guru dan 28 siswa SMP; penilaian berbicara otentik	Sikap, pengetahuan, dan keterampilan dinilai melalui teknik berbeda; penilaian otentik dinilai layak untuk berbicara
Al Ayubi et al. (2021)	Studi kasus dua guru bahasa Inggris di madrasah aliyah; persepsi terhadap penilaian otentik	Guru mendukung dan menilainya bermanfaat, tetapi kesulitan mengelola kegiatan kelas, menjaga konsistensi penskoran, dan memastikan keaslian karya

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Penulis dan tahun	Konteks dan fokus	Temuan utama
Nurkumaila et al. (2025)	Telaah sistematis praktik penilaian berbicara EFL (2015–2025)	Pendekatan berbasis kinerja dan teknologi dominan; bias penilai, literasi asesmen, dan kecemasan menjadi tantangan
Sukmawati et al. (2026)	Studi kasus guru SMA; strategi penilaian berbicara otentik	Tugas kehidupan nyata dan rubrik analitik dipakai; kendala waktu, ukuran kelas, pelatihan, dan panduan kurikulum

Beberapa pola dapat dicermati dari sintesis tersebut. Pertama, terdapat konsistensi arah: penilaian otentik diterima dan dianggap layak oleh guru, serta dipraktikkan melalui tugas kehidupan nyata dan rubrik analitik. Kedua, hambatan yang dilaporkan berulang pada tiga ranah, yaitu sumber daya (waktu dan jumlah siswa), kapasitas (literasi asesmen dan validasi rubrik), serta mutu pengukuran (konsistensi dan bias penilai). Ketiga, bukti tentang dampak penilaian otentik terhadap peningkatan keterampilan komunikasi secara terukur masih terbatas karena banyak penelitian berupa studi persepsi atau studi kasus kualitatif; kajian sistematis bahkan merekomendasikan penelitian longitudinal untuk menguji pengaruh kerangka asesmen terpadu terhadap peningkatan berbicara, motivasi, dan partisipasi (Nurkumaila et al., 2025). Keempat, sebagian besar penelitian berskala kecil, misalnya satu guru pada satu kelas sehingga tidak dapat digeneralisasi (Sukmawati et al., 2026). Dengan demikian klaim bahwa penilaian otentik meningkatkan keterampilan komunikasi pada artikel ini bersandar terutama pada penalaran teoretis dan bukti tidak langsung, dan perlu diuji lebih jauh.

Tantangan dan Strategi Mitigasi

Berdasarkan sintesis, enam tantangan utama dapat diidentifikasi beserta strategi yang dapat ditempuh.

Tantangan pertama adalah keterbatasan waktu dan jumlah siswa. Strategi yang relevan adalah penilaian bertahap sepanjang semester, penggunaan kerja berpasangan atau kelompok kecil, penilaian di luar kelas melalui rekaman, serta pemilihan sampel kinerja pada setiap periode agar semua siswa terwakili dalam beberapa kali penilaian. Kunci strategi ini adalah menjadikan penilaian bagian dari pembelajaran sehari-hari, sehingga tidak menambah beban terpisah.

Tantangan kedua adalah literasi asesmen guru yang belum merata. Strategi yang dapat ditempuh adalah pelatihan berbasis praktik pada forum guru mata pelajaran, pengembangan rubrik bersama, dan pendampingan kepala sekolah atau pengawas. Pembelajaran profesional yang berpusat pada pengalaman menyusun, menggunakan, dan merevisi rubrik lebih berpotensi mengubah praktik dibandingkan pelatihan yang bersifat ceramah (Popham, 2009).

Tantangan ketiga adalah reliabilitas dan bias penilai. Strategi yang disarankan mencakup penggunaan rubrik analitik berdeskriptor perilaku, pelatihan penilai dengan sampel kinerja, moderasi antarpenilai, dan penggunaan lebih dari satu

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

kesempatan penilaian sehingga keputusan tidak bergantung pada satu penampilan (Jonsson & Svingby, 2007; Fulcher, 2010). Penilaian sejawat dan diri dapat melengkapi, bukan menggantikan, penilaian guru.

Tantangan keempat adalah kecemasan dan kesenjangan partisipasi siswa. Strategi yang dapat dipakai antara lain tugas berpasangan yang membangun rasa aman, latihan bertahap sebelum penilaian bernilai, kriteria yang jelas, dan umpan balik yang menekankan perkembangan. Guru perlu memantau agar siswa yang lebih percaya diri tidak mendominasi pasangan, misalnya dengan menetapkan peran bergiliran (Horwitz et al., 1986; Sukmawati et al., 2026).

Tantangan kelima adalah pemanfaatan teknologi yang belum optimal. Teknologi tersedia namun jarang dipakai oleh seorang guru yang diteliti, antara lain karena keterampilan digital yang terbatas (Sukmawati et al., 2026), padahal asesmen yang didukung teknologi dilaporkan dapat memperbaiki umpan balik dan keterlibatan (Nurkumaila et al., 2025). Perekaman audio dan video sederhana, portofolio digital, serta platform umpan balik dapat mengurangi beban waktu. Namun penggunaannya harus memperhatikan privasi siswa, keterjangkauan perangkat, dan kemungkinan bantuan kecerdasan buatan yang mengurangi keaslian kinerja. Al Ayubi et al. (2021) mencatat persoalan keaslian karya sebagai kesulitan yang dirasakan guru, sehingga tugas yang menuntut kinerja langsung dan refleksi proses lebih tahan terhadap persoalan tersebut dibandingkan tugas yang hanya mengumpulkan produk.

Tantangan keenam adalah ketidakjelasan panduan operasional. Kurikulum yang memberi keleluasaan tetapi minim panduan teknis menimbulkan variasi praktik antarguru (Sukmawati et al., 2026). Strategi yang dibutuhkan adalah penyediaan contoh perangkat asesmen, rubrik acuan, dan portofolio praktik baik pada tingkat sekolah dan dinas pendidikan, yang tetap memberi ruang adaptasi.

Model Konseptual dan Proposisi

Sintesis di atas memungkinkan perumusan model konseptual manajemen penilaian otentik untuk mendorong keterampilan komunikasi. Model ini memandang manajemen penilaian sebagai siklus empat fungsi: perencanaan, pengorganisasian, pelaksanaan, dan pengendalian mutu, yang berulang pada setiap unit pembelajaran. Siklus tersebut dipengaruhi oleh faktor pendukung berupa literasi asesmen guru, kepemimpinan kepala sekolah, sumber daya waktu dan teknologi, serta kebijakan kurikulum. Keluaran langsungnya berupa penilaian yang otentik, valid, dan andal, sedangkan keluaran lanjutannya berupa peningkatan keterampilan komunikasi melalui tiga mekanisme: peluang interaksi, umpan balik formatif, dan washback positif. Kinerja yang terukur kembali menjadi masukan bagi perencanaan berikutnya sehingga siklus berlangsung berkelanjutan.

Berdasarkan model tersebut, lima proposisi dirumuskan untuk diuji pada penelitian selanjutnya.

a. Semakin tinggi keotentikan tugas penilaian, yang diukur melalui dimensi tugas, konteks, bentuk kinerja, kriteria, dan keterlibatan siswa, semakin tinggi keterampilan komunikasi siswa.
b. Kualitas manajemen penilaian, yang mencakup perencanaan, pengorganisasian, pelaksanaan, dan pengendalian mutu, memoderasi pengaruh penilaian otentik

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

terhadap keterampilan komunikasi; penilaian otentik yang dikelola dengan baik menghasilkan peningkatan yang lebih besar.

c. Frekuensi dan kualitas umpan balik formatif memediasi hubungan antara penilaian otentik dan keterampilan komunikasi.
d. Literasi asesmen guru dan penggunaan rubrik yang tervalidasi berhubungan positif dengan reliabilitas penskoran dan kualitas umpan balik.
e. Dukungan kepala sekolah berupa waktu, pelatihan, dan kolaborasi moderasi memperkuat pengaruh manajemen penilaian otentik terhadap keterampilan komunikasi, sedangkan ukuran kelas yang besar melemahkannya apabila tidak diimbangi strategi pengorganisasian.

Implikasi

Bagi guru, kajian ini mengimplikasikan perlunya memperlakukan penilaian sebagai proses yang dikelola, bukan hanya kumpulan tugas. Guru dianjurkan memulai dengan satu atau dua tugas komunikatif per unit, menggunakan rubrik analitik yang ringkas, memberikan umpan balik tertulis atau lisan yang spesifik, dan mengumpulkan sampel kinerja dalam portofolio. Bagi kepala sekolah, kajian ini menekankan pentingnya menyediakan waktu moderasi, memfasilitasi pelatihan literasi asesmen, dan menciptakan budaya saling berbagi perangkat penilaian. Bagi pembuat kebijakan, kajian ini menyarankan penyediaan panduan operasional dan contoh perangkat asesmen yang tetap memberi keleluasaan adaptasi. Bagi peneliti, proposisi yang dirumuskan dapat diuji melalui desain kuasi-eksperimen, studi kasus ganda, penelitian tindakan kelas, atau desain campuran dengan pengukuran keterampilan komunikasi yang valid.

Implikasi lain berkaitan dengan pemantauan keberhasilan manajemen penilaian. Sekolah dapat menetapkan indikator proses yang sederhana, misalnya proporsi unit pembelajaran yang memuat sedikitnya satu tugas komunikatif dengan rubrik, jumlah sesi moderasi per semester, ketepatan waktu penyampaian umpan balik kepada siswa, serta persentase siswa yang memperoleh kesempatan tampil lebih dari satu kali. Indikator keluaran dapat berupa perkembangan skor pada dimensi rubrik dari waktu ke waktu serta refleksi siswa tentang kepercayaan diri berkomunikasi. Pemantauan semacam ini menjadikan manajemen penilaian dapat dievaluasi dan diperbaiki, sekaligus menyediakan data bagi penelitian lanjutan.

Keterbatasan Kajian

Kajian ini memiliki beberapa keterbatasan. Pertama, sebagai telaah pustaka, artikel ini tidak mengumpulkan data primer sehingga tidak dapat memastikan besaran pengaruh penilaian otentik terhadap keterampilan komunikasi. Kedua, bukti empiris yang dirujuk didominasi studi kasus dan studi persepsi pada konteks tertentu, sehingga generalisasi harus dilakukan dengan hati-hati. Ketiga, penelusuran terbatas pada sumber yang dapat diakses penulis dan sebagian artikel empiris hanya dapat dibaca pada bagian abstrak dan bagian yang tersedia, sehingga kemungkinan terdapat studi relevan yang belum terjangkau atau rincian yang belum tertelaah. Keempat, artikel ini berfokus pada keterampilan berbicara sebagai wujud utama

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

komunikasi lisan, sehingga penerapan pada keterampilan menulis dan menyimak memerlukan kajian tersendiri.

KESIMPULAN DAN SARAN

Berdasarkan telaah pustaka, dapat disimpulkan bahwa penilaian otentik merupakan pendekatan yang selaras dengan hakikat keterampilan komunikasi karena menempatkan siswa pada tugas bermakna yang menuntut penggunaan bahasa secara utuh, dengan kriteria yang transparan dan umpan balik yang berkelanjutan. Penilaian otentik mendorong keterampilan komunikasi melalui peluang berinteraksi, penguatan umpan balik formatif, dan pengaruh balik positif terhadap pembelajaran. Keberhasilannya sangat bergantung pada manajemen yang mencakup perencanaan tugas dan rubrik, pengorganisasian waktu dan sumber daya, pelaksanaan yang bertahap dan suportif, serta pengendalian mutu melalui moderasi dan tindak lanjut. Hambatan berupa keterbatasan waktu dan jumlah siswa, literasi asesmen, reliabilitas penskoran, kecemasan siswa, pemanfaatan teknologi, dan panduan operasional dapat dimitigasi melalui strategi yang dirumuskan. Model siklus manajemen dan lima proposisi menjadi landasan bagi penelitian empiris.

Saran dirumuskan untuk tiga pihak. Bagi guru bahasa Inggris, disarankan menerapkan penilaian otentik secara bertahap dengan rubrik yang divalidasi bersama rekan sejawat dan memanfaatkan penilaian diri serta sejawat. Bagi sekolah dan pembuat kebijakan, disarankan menyediakan pelatihan literasi asesmen, waktu moderasi, serta panduan dan contoh perangkat yang operasional. Bagi peneliti, disarankan menguji proposisi melalui penelitian longitudinal atau kuasi-eksperimen yang mengukur keterampilan komunikasi dengan instrumen valid, serta memperluas kajian pada keterampilan bahasa lain dan pada konteks jenjang yang berbeda.

DAFTAR PUSTAKA

Al Ayubi, S., Erlina, D., & Desvitasari, D. (2021). English teachers' perception on the use of authentic assessment in EFL classroom at MA Al-Fatah Palembang. Jadila: Journal of Development and Innovation in Language and Literature Education, 2(1), 73–87. https://doi.org/10.52690/jadila.v2i1.185

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. AERA.

Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford University Press.

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139–148.

Boud, D., & Falchikov, N. (2007). Rethinking assessment in higher education: Learning for the longer term. Routledge.

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Brookhart, S. M. (2013). How to create and use rubrics for formative assessment and grading. ASCD.

Brown, H. D., & Abeywickrama, P. (2019). Language assessment: Principles and classroom practices (3rd ed.). Pearson.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), Language and communication (pp. 2–27). Longman.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1(1), 1–47.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. Issues in Applied Linguistics, 6(2), 5–35.

Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Council of Europe Publishing.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.

Ellis, R. (2003). Task-based language learning and teaching. Oxford University Press.

Fulcher, G. (2010). Practical language testing. Hodder Education.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52(3), 67–86.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. School Effectiveness and School Improvement, 9(2), 157–191.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. The Modern Language Journal, 70(2), 125–132.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), Sociolinguistics (pp. 269–293). Penguin.

Inayah, N., Komariah, E., & Nasir, A. (2019). The practice of authentic assessment in an EFL speaking classroom. Studies in English Language and Education, 6(1), 152–162. https://doi.org/10.24815/siele.v6i1.13069

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. Educational Research Review, 2(2), 130–144.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 17–64). American Council on Education/Praeger.

Kunandar. (2013). Penilaian autentik (penilaian hasil belajar peserta didik berdasarkan Kurikulum 2013). Rajawali Pers.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE Publications.

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), Handbook of second language acquisition (pp. 413–468). Academic Press.

Luoma, S. (2004). Assessing speaking. Cambridge University Press.

Majid, A. (2014). Penilaian autentik proses dan hasil belajar. PT Remaja Rosdakarya.

Messick, S. (1996). Validity and washback in language testing. Language Testing, 13(3), 241–256.

Moleong, L. J. (2017). Metodologi penelitian kualitatif (Edisi revisi). PT Remaja Rosdakarya.

Nunan, D. (2004). Task-based language teaching. Cambridge University Press.

Nurkumaila, F., Salma, Z., Husnullah, A. H., Dewi, R. S., & Maulidya, R. A. (2025). Systematic literature review on speaking assessment practices in EFL contexts. Journal of English Teaching, Linguistics, Literature, and Education (JETLEE). https://journal.uinsuna.ac.id/index.php/jetlee/article/view/7247

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. Educational Research Review, 9, 129–144.

Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? Educational Leadership, 48(5), 60–63.

Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 21 Tahun 2022 tentang Standar Penilaian Pendidikan pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? Theory Into Practice, 48(1), 4–11.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18(2), 119–144.

Shepard, L. A. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 4–14.

Shohamy, E. (2001). The power of tests: A critical perspective on the uses of language tests. Longman.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339.

Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72(7), 534–539.

Sukmawati, A. A. A. M., Artini, L. P., Suarcaya, P., & Marsakawati, N. P. E. (2026). Exploring authentic assessment strategies for EFL speaking: The experience of an Indonesian high school teacher at senior high school. IDEAS: Journal on Language Teaching and Learning, Linguistics and Literature, 14(1), 1897–1916. https://doi.org/10.24256/ideas.v14i1.10408

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), Input in second language acquisition (pp. 235–253). Newbury House.

Terry, G. R., & Rue, L. W. (2010). Dasar-dasar manajemen (G. A. Ticoalu, Penerj.). Bumi Aksara.

Topping, K. (1998). Peer assessment between students in colleges and universities. Review of Educational Research, 68(3), 249–276.

Vol 5 No 2 (2026) 318 – 338 E-ISSN 2962-2840
DOI: 10.56672/attadris.v5i2.648

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management, 14(3), 207–222.

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.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wiggins, G. (1990). The case for authentic assessment. Practical Assessment, Research & Evaluation, 2(2).

Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. Jossey-Bass.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70.