JPFS 9 (2) 2026 393-413

Jurnal Pendidikan Fisika dan Sains (JPFS)

Journal homepage: http:// journal.unucirebon.ac.id/index.php/jpfs

Persepsi Kebermanfaatan Canvas Asesmen dan Kualitas Rancangan Penilaian Guru pada Tahapan PGBU dan Praktik TEL

Iin Suminar*[1], Nurdini[2] ,Utami Widyaiswari[3], Lathifa Nur Ramdhania[4], M. Reza Dwi Saputra[5], Intan Indah Megasari[6]

[1,2,5]Pendidikan Fisika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia

[3,4]Fisika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia

[6]Pendidikan Pancasila dan Kewarganegaraan, Fakultas Pendidikan Ilmu Pengetahuan Sosial, Universitas Pendidikan Indonesia

*Email: iinsuminar28@upi.edu

DOI: https://doi.org/10.52188/jpfs.v9i2.2588

Accepted: 27 Oktober 2026 Approved: 20 November 2026 Published: 23 November 2026

ABSTRAK

Penelitian ini bertujuan menganalisis kualitas rancangan penilaian guru menggunakan Canvas Asesmen Berbasis Proyek, persepsi kebermanfaatan (*perceived usefulness*) canvas, dan hubungan keduanya. Canvas memetakan aktivitas dan bukti belajar pada tahapan Pikir–Gambar–Buat–Uji (PGBU), yaitu alur pengembangan solusi rekayasa dalam pembelajaran berbasis proyek serta menghubungkannya dengan dengan instrumen penilaian dan praktik *Technology and Engineering Literacy (TEL)*, yaitu kapasitas menggunakan, memahami, dan mengevaluasi teknologi serta mengembangkan solusi untuk mencapai tujuan. Penelitian kuantitatif deskriptif-evaluatif ini menggabungkan analisis dokumen dan survei potong lintang terhadap 44 guru. Rancangan dinilai oleh satu penilai menggunakan rubrik sembilan dimensi berskala 1–4; persepsi guru diukur dengan angket Likert 11 butir dan dianalisis menggunakan *Rasch Rating Scale Model*; hubungan keduanya diuji dengan korelasi Spearman. Rerata kualitas rancangan 3,52 pada skala 4; dengan 81,8% rancangan berkategori sangat baik. Keselarasan bukti belajar dengan metode atau instrumen penilaian memperoleh rerata terendah (3,07). Sebanyak 90,3% respons berada pada kategori setuju dan sangat setuju. Reliabilitas responden 0,816 dan separasi 2,105, tetapi penargetan butir belum optimal. Korelasi persepsi kebermanfaatan dan kualitas rancangan positif, sangat lemah, dan tidak signifikan ($r_s = 0{,}087;\ p = 0{,}576$). Persepsi positif tidak otomatis menghasilkan rancangan berkualitas sehingga pendampingan diperlukan untuk pengembangan instrumen, bukti individual, dan umpan balik.

Kata kunci: Asesmen Berbasis Proyek; Bukti Belajar; PGBU; Model Rasch; Literasi Teknologi Dan Rekayasa

ABSTRACT

This study examined the quality of teachers' assessment designs developed with a project-based assessment canvas, teachers' perceived usefulness of the canvas, and their relationship. The canvas maps activities and learning evidence across PGBU (Pikir–Gambar–Buat–Uji; Think–Draw–Make–Test), an engineering-design workflow in project-based learning, and links them with assessment instruments and Technology and Engineering Literacy (TEL) practices, the capacity to use, understand, and evaluate technology and to develop solutions to achieve goals.. This descriptive-evaluative quantitative study combined document analysis and a cross-sectional survey of 44 teachers. A single rater scored the designs using a nine-dimension, four-point rubric; an 11-item Likert questionnaire was

analyzed with the Rasch Rating Scale Model; Spearman's correlation tested the relationship. The mean design score was 3.52 out of 4, with 81.8% of designs rated very good. Alignment of learning evidence with assessment methods or instruments had the lowest mean (3.07). Agree and strongly agree responses reached 90.3%. Rasch person reliability was 0.816 and separation 2.105, although item targeting was suboptimal. Perceived usefulness and design quality showed a very weak, positive, nonsignificant correlation *($r_s = 0.087,\ p = 0.576$)*. Positive perceptions did not ensure high-quality designs; support is needed to develop instruments, individual evidence, and feedback.

Keyword: project-based assessment; learning evidence; PGBU; Rasch model; technology and engineering literacy

@2026 Pendidikan Fisika FKIP Universitas Nahdlatul Ulama Cirebon

PENDAHULUAN

Produk nyata sering menjadi bagian paling terlihat dalam *project-based learning* (PjBL). Namun, persoalan pendidikan yang lebih luas bukan sekadar kemampuan peserta didik menghasilkan artefak, melainkan kesiapan mereka menyelidiki masalah autentik, menggunakan informasi secara kritis, menerapkan pengetahuan, berkolaborasi, dan mengembangkan solusi yang dapat dipertanggungjawabkan. Dalam PjBL, produk, performa, atau solusi merupakan representasi proses *inquiry*, bukan satu-satunya tujuan pembelajaran. Kajian mutakhir menunjukkan bahwa PjBL dapat mendukung hasil akademik, keterampilan interpersonal, pemecahan masalah, dan penerapan pengetahuan apabila proyek berangkat dari masalah bermakna dan penyelidikannya terstruktur (Crespí et al., 2022; Markula & Aksela, 2022; Suyantiningsih et al., 2023; Zhang & Ma, 2023; Lavado-Anguera et al., 2024).

Proses belajar dalam PjBL berlangsung sepanjang tahapan proyek. Peserta didik dituntut untuk memahami dan merumuskan masalah, mengajukan pertanyaan, mencari dan menyeleksi informasi, mengembangkan gagasan, membuat rancangan, membuat solusi, melakukan pengujian, menerima umpan balik, memperbaiki hasil, serta merefleksikan proses yang telah dilakukan. PjBL merupakan pembelajaran yang bersifat dinamis, kolaboratif, dan iteratif, sehingga guru seyogianya dapat memfasilitasi *inquiry*, membangun budaya kolaborasi, menjaga keterhubungan kegiatan dengan tujuan belajar, serta menyediakan ruang untuk refleksi dan revisi (Markula & Aksela, 2022; Pupik Dean et al., 2023; Halawa et al., 2024). Dengan demikian, pengalaman belajar peserta didik tersebar pada berbagai keputusan, tindakan, dan perubahan yang terjadi selama proyek berlangsung.

Karakteristik tersebut membawa konsekuensi penting bagi perancangan asesmen. Jika penilaian hanya diarahkan pada produk akhir, bukti mengenai cara peserta didik memahami masalah, menerapkan konsep, membuat keputusan, berkolaborasi, mengatasi kegagalan, dan memperbaiki solusi dapat terabaikan. Sebaliknya, penilaian proses tanpa memperhatikan kualitas solusi juga belum menggambarkan ketercapaian tujuan proyek secara utuh. Asesmen berbasis proyek perlu menangkap proses belajar dan produk atau performa secara terpadu, dengan keselarasan eksplisit antara tujuan, aktivitas, bukti, dan keputusan penilaian (Harris et al., 2022; Zhao et al., 2023; van der Steen et al., 2023).

Bukti belajar dalam proyek dapat berupa rumusan masalah, catatan penyelidikan, argumentasi pemilihan solusi, gambar rancangan, pembagian peran, produk antara, dokumentasi pembuatan, data pengujian, revisi, presentasi, dan refleksi. Sebagian bukti digunakan secara formatif untuk mengetahui posisi belajar peserta didik dan menentukan dukungan berikutnya. Asesmen formatif membentuk siklus bukti–umpan balik–refleksi–perbaikan melalui pertanyaan guru, pemeriksaan produk antara, observasi, penilaian diri atau teman sebaya, diskusi kemajuan, serta revisi rancangan. Bukti mutakhir menunjukkan bahwa asesmen formatif bermanfaat bagi pembelajaran, tetapi dampaknya bergantung pada kualitas rancangan dan tindak lanjutnya (Abd Halim et al., 2024; Sortwell et al., 2024; Yao et al., 2024). Selain itu, pelibatan peserta didik melalui refleksi, penilaian diri, dan penilaian teman sebaya juga membutuhkan kriteria keberhasilan serta kesempatan untuk menggunakan umpan balik, bukan sekadar menerimanya (Zulliger et al., 2022; de Kleijn, 2023; Woitt et al., 2025). Karena itu, desain umpan balik harus mempertimbangkan pesan, pelaksana, konteks, dan tindakan lanjutan secara terpadu (Panadero & Lipnevich, 2022).

Merancang asesmen semacam itu bukan tugas sederhana. Guru harus menerjemahkan tujuan pembelajaran menjadi bukti yang dapat diamati, memilih teknik dan instrumen, menetapkan kriteria kualitas, membedakan fungsi formatif dan sumatif, serta menghubungkan proses dengan produk. Literasi asesmen guru mencakup pemahaman konseptual, keterampilan praktik, dan pertimbangan sosioemosional dalam menggunakan bukti untuk mendukung pembelajaran (Yan & Pastore, 2022). Kajian terkini menunjukkan bahwa guru masih menghadapi kesenjangan antara pemahaman asesmen dan penerapannya dalam perencanaan yang kontekstual, sehingga penguatan literasi dan pendampingan tetap diperlukan (Pastore, 2023; Saky et al., 2025; Lei & Lei, 2026). Tantangan ini semakin besar dalam PjBL karena bukti belajar tersebar di berbagai tahap dan berbentuk pengetahuan, keterampilan, kolaborasi, serta produk.

Kebutuhan untuk mengorganisasi proses dan bukti belajar menjadi semakin penting pada proyek yang melibatkan praktik rekayasa. Dalam penelitian ini, Pikir–Gambar–Buat–Uji (PGBU) digunakan sebagai alur pengembangan solusi. PGBU bukan asesmen dan bukan instrumen penilaian, melainkan alur kerja rekayasa yang membantu mengorganisasi aktivitas peserta didik: menganalisis masalah dan mengembangkan gagasan pada tahap Pikir, merepresentasikan solusi pada tahap Gambar, mewujudkan rancangan pada tahap Buat, serta menguji fungsi dan kualitas solusi pada tahap Uji. Penerapan PGBU dalam proyek autentik menunjukkan potensinya untuk menghubungkan konsep, desain, pembuatan, pengujian, dan pemecahan masalah (Suminar et al., 2025). PGBU bersifat iteratif dan selaras dengan praktik desain rekayasa yang menghubungkan perancangan, pembuatan, pengujian, evaluasi, dan perbaikan (Halawa et al., 2024; Lavado-Anguera et al., 2024; Kidd et al., 2026). PGBU dengan demikian berfungsi sebagai alur operasional di dalam PjBL: PjBL menyediakan kerangka pembelajaran berbasis masalah autentik dan *inquiry*, sedangkan PGBU menata pengembangan solusi secara terstruktur dan iteratif.

Setiap tahap PGBU dapat menghasilkan bukti belajar yang berbeda. Tahap Pikir dapat menghasilkan identifikasi masalah, pertanyaan, informasi pendukung, penerapan konsep, alternatif solusi, dan kriteria keberhasilan. Tahap Gambar dapat menghasilkan sketsa, diagram, spesifikasi, pemilihan bahan, dan argumentasi desain. Tahap Buat dapat memberikan bukti keterampilan menggunakan alat dan bahan, penerapan rancangan, kolaborasi, pemecahan masalah, dan dokumentasi perubahan. Tahap Uji dapat menghasilkan data fungsi produk, kesesuaian terhadap kriteria, interpretasi hasil, identifikasi kelemahan, dan rencana perbaikan. Bukti dari setiap tahap memungkinkan asesmen menangkap inquiry sekaligus kualitas solusi.

PGBU dipilih sebagai kerangka pemetaan bukti belajar karena tiga alasan. Pertama, setiap tahap PGBU menghasilkan keluaran yang berbeda dan dapat diamati, sehingga batas antartahap dapat dijadikan titik pengumpulan bukti dan pemberian umpan balik yang jelas bagi guru. Kedua, empat tahap PGBU menyederhanakan siklus desain rekayasa, yaitu mendefinisikan masalah, merancang, membuat, menguji, dan memperbaiki, ke dalam istilah yang ringkas dan akrab bagi guru, sehingga dapat digunakan lintas mata pelajaran tanpa menghilangkan sifat iteratifnya (Halawa et al., 2024; Kidd et al., 2026). Ketiga, PGBU telah diterapkan dalam proyek autentik di sekolah menengah (Suminar et al., 2025), tetapi pemanfaatannya masih berfokus pada aktivitas dan produk peserta didik, belum pada perencanaan asesmen oleh guru. Dengan menempatkan bukti belajar pada setiap tahap PGBU, asesmen dapat diarahkan pada proses berpikir, keputusan desain, dan perbaikan solusi, bukan hanya pada produk akhir.

Pemetaan bukti belajar pada tahapan PGBU berkaitan dengan kerangka *Technology and Engineering Literacy* (TEL), khususnya area *Design and Systems*. Kerangka TEL memuat tiga praktik lintas area, yaitu *Understanding Technological Principles, Developing Solutions and Achieving Goals*, serta *Communicating and Collaborating* (National Assessment Governing Board, 2018). Tahap Pikir dan Gambar dapat mengungkap pemahaman prinsip teknologi dan pengembangan rancangan solusi; tahap Buat dan Uji memberi peluang untuk menilai pewujudan solusi, penggunaan sumber daya, pengujian tujuan, dan perbaikan. Praktik komunikasi dan kolaborasi dapat muncul di seluruh tahap melalui diskusi, pembagian peran, dokumentasi, argumentasi keputusan, dan penyampaian hasil.

Dengan demikian, PjBL, PGBU, dan TEL tidak diposisikan sebagai tiga pendekatan yang berdiri sendiri, melainkan sebagai tiga lapis yang saling melengkapi dalam satu rancangan. PjBL merupakan kerangka pedagogis yang menyediakan masalah autentik, *inquiry*, kolaborasi, dan produk sebagai konteks belajar. PGBU merupakan alur kerja di dalam proyek yang menata kapan dan melalui aktivitas apa bukti belajar dihasilkan. TEL merupakan literasi teknologi dan rekayasa yang dirumuskan

dalam kerangka NAEP dan menjadi acuan untuk menafsirkan bukti tersebut melalui tiga praktiknya, yaitu memahami prinsip teknologi, mengembangkan solusi dan mencapai tujuan, serta berkomunikasi dan berkolaborasi. Asesmen menjadi penghubung ketiganya: bukti yang muncul pada tahap PGBU dalam konteks PjBL dinilai menggunakan metode dan kriteria yang merujuk pada praktik TEL.

Persoalannya, pemetaan tersebut menuntut guru mengoordinasikan kompetensi, aktivitas setiap tahap PGBU, bukti yang diharapkan, waktu pengumpulan, metode dan pelaksana penilaian, kriteria keberhasilan, serta tindak lanjut. Tanpa alat bantu perencanaan yang terstruktur, komponen tersebut dapat disusun terpisah sehingga aktivitas, bukti, dan instrumen tidak sepenuhnya selaras. Studi desain asesmen formatif menunjukkan bahwa ketidakselarasan tujuan, aktivitas, bukti, dan tindak lanjut masih ditemukan bahkan pada rancangan guru berpengalaman (van der Steen et al., 2023). Pengembangan asesmen yang terarah juga memerlukan dukungan profesional agar guru dapat menghubungkan bukti dengan perkembangan belajar dan langkah berikutnya (Baartman & Gulikers, 2025).

Canvas Asesmen Berbasis Proyek dikembangkan sebagai alat bantu visual untuk menjawab kebutuhan tersebut. Canvas memandu guru memetakan konteks proyek, tahapan PGBU, aktivitas peserta didik, bukti belajar, instrumen penilaian, praktik TEL, serta umpan balik dan revisi dalam satu tampilan.. Canvas asesmen ini berfungsi dalam menopang proses berpikir agar keterhubungan antarkomponen dapat diperiksa sebelum pembelajaran. Prinsip ini sejalan dengan temuan bahwa alat desain pembelajaran perlu menjaga keselarasan antara maksud pedagogis dan fitur alat agar benar-benar mendukung inquiry dan pembelajaran aktif (Børte & Lillejord, 2024).

Canvas ini berbeda dari alat bantu perencanaan asesmen yang telah dikenal. Kisi-kisi atau tabel spesifikasi umumnya memetakan tujuan ke butir soal sehingga berorientasi pada tes, sedangkan rubrik proyek umumnya digunakan untuk menilai kinerja setelah produk selesai. Kerangka *Assessment Design Decisions* membantu pendidik mengambil keputusan mengenai tujuan, konteks, tugas, interaksi, umpan balik, dan capaian belajar, tetapi dikembangkan untuk perancangan asesmen di pendidikan tinggi dan tidak menautkan keputusan tersebut pada tahap pengerjaan proyek (Bearman et al., 2016). Alat desain pembelajaran menyelaraskan pedagogi dan teknologi secara umum (Børte & Lillejord, 2024), sedangkan kajian desain asesmen formatif dan progresi belajar menekankan hubungan bukti dengan langkah pembelajaran berikutnya tanpa struktur tahapan rekayasa (van der Steen et al., 2023; Baartman & Gulikers, 2025). Keunggulan Canvas Asesmen Berbasis Proyek terletak pada pemetaan berjenjang dalam satu tampilan: setiap tahap PGBU dihubungkan dengan aktivitas, bukti proses atau produk, instrumen penilaian, dan praktik TEL, serta dengan alur umpan balik dan revisi. Struktur ini memungkinkan guru dan penelaah memeriksa keselarasan antarkomponen sebelum pembelajaran dilaksanakan.

Penelitian terdahulu memberikan dasar penting, tetapi masih menyisakan tiga kesenjangan. Pertama, kajian PjBL dan praktik desain rekayasa umumnya berfokus pada karakteristik implementasi dan dampaknya terhadap hasil belajar peserta didik (Markula & Aksela, 2022; Zhang & Ma, 2023; Kidd et al., 2026), sedangkan cara guru merencanakan pengumpulan bukti pada setiap tahap pengembangan solusi belum banyak dikaji. Kedua, kajian literasi asesmen menunjukkan adanya kesenjangan antara pemahaman dan praktik perencanaan asesmen guru (Pastore, 2023; Lei & Lei, 2026), tetapi belum menyediakan alat bantu yang secara khusus menautkan tahapan rekayasa dengan bukti belajar, metode penilaian, dan praktik TEL. Ketiga, evaluasi alat bantu perencanaan kerap dilakukan melalui persepsi pengguna, padahal penilaian diri tidak selalu sejalan dengan ukuran kinerja yang objektif (Ernst et al., 2023). Akibatnya, belum diketahui apakah persepsi positif guru terhadap suatu alat bantu perencanaan sejalan dengan kualitas rancangan yang benar-benar dihasilkan. Penelitian ini menjawab kesenjangan tersebut dengan menilai secara bersamaan kualitas rancangan guru dan persepsi kebermanfaatan canvas pada responden yang sama.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan menganalisis kualitas rancangan penilaian yang disusun guru menggunakan Canvas Asesmen Berbasis Proyek, persepsi kebermanfaatan canvas menurut guru, serta hubungan keduanya. Secara khusus, penelitian ini menjawab tiga pertanyaan: (1) bagaimana kualitas rancangan penilaian yang disusun guru menggunakan canvas pada setiap dimensi; (2) bagaimana persepsi kebermanfaatan canvas menurut guru berdasarkan analisis Rasch; dan (3) apakah terdapat hubungan antara persepsi kebermanfaatan dan kualitas rancangan. Penelitian dilakukan pada tahap perancangan dan belum mencakup implementasi canvas dalam pembelajaran. Karena itu, penelitian tidak menyimpulkan efektivitas canvas terhadap hasil belajar peserta didik. Kontribusinya berupa bukti awal mengenai sejauh mana struktur canvas dipersepsikan

bermanfaat, dapat digunakan guru untuk memetakan bukti belajar, dan menghasilkan rancangan penilaian yang berkualitas. Temuan diharapkan menjadi dasar untuk menyempurnakan canvas, merancang pendampingan guru, dan menguji implementasinya di kelas.

METODE

Desain Penelitian

Penelitian ini menggunakan desain kuantitatif deskriptif-evaluatif dengan analisis dokumen dan survei. Analisis dokumen digunakan untuk menilai kualitas rancangan canvas yang disusun guru, sedangkan survei dilakukan satu kali setelah guru menggunakan canvas untuk mengukur persepsi mereka terhadap kebermanfaatan canvas. Kedua sumber data dianalisis secara terpisah, kemudian dipertemukan pada tingkat responden untuk memperoleh gambaran mengenai kualitas rancangan yang dihasilkan dan persepsi guru setelah menggunakan canvas. Desain ini sesuai untuk mengevaluasi suatu alat bantu perencanaan pada tahap penggunaan awal tanpa menguji pengaruhnya terhadap hasil belajar peserta didik (Creswell & Creswell, 2023).

Partisipan dan Penetapan Sampel

Partisipan penelitian merupakan guru yang mengikuti kegiatan pengembangan profesional tentang perancangan asesmen berbasis proyek. Data dikumpulkan melalui pengisian angket dan penyerahan rancangan penilaian canvas pada 1–13 Agustus 2026. Sampel analisis ditentukan menggunakan *criterion-based sampling*, yaitu seluruh data yang memenuhi kriteria kelengkapan, keunikan rancangan penilaian, keterverifikasian identitas, dan keterpasangan antara angket dan rancangan penilaian.

Kriteria inklusi penelitian meliputi: (1) responden berprofesi sebagai guru; (2) mengisi seluruh butir angket persepsi; (3) menyerahkan satu rancangan penilaian canvas yang dapat dibuka dan dibaca; (4) rancangan penilaian memuat dua bagian utama template, termasuk pemetaan bukti belajar dan praktik TEL; (5) rancangan penilaian merupakan karya yang unik; dan (6) identitas pemilik rancangan penilaian dapat dipasangkan dengan data angket. Berkas satu halaman atau tidak lengkap dikeluarkan karena tidak memuat keseluruhan struktur canvas yang diperlukan untuk menilai pemetaan bukti belajar pada tahapan PGBU dan praktik TEL. Proses penetapan sampel ditunjukkan pada Tabel 1.

Tabel 1. Proses Penetapan Sampel Penelitian

Tahap penyaringan	Jumlah	Keputusan
Respons awal yang terkumpul	82	Data awal
Respons unik dan lengkap setelah pembersihan identitas dan duplikasi respons	64	Memenuhi kelengkapan angket
Data guru dengan produk yang dapat dipasangkan secara sementara	56	Memasuki audit produk (8 respons tanpa produk yang dapat dipasangkan)
Data dengan dokumen tersedia namun identitas produk tidak terkonfirmasi	3	Dikeluarkan
Produk satu halaman atau tidak memuat struktur canvas secara lengkap	7	Dikeluarkan
Produk identik yang diserahkan oleh dua guru	2	Keduanya dikeluarkan
Sampel akhir berpasangan	44	Dianalisis

Dari 56 data yang memasuki audit produk, 12 data dikeluarkan. Dengan demikian, sampel akhir yang dianalisis berjumlah 44 pasangan data angket dan rancangan penilaian.

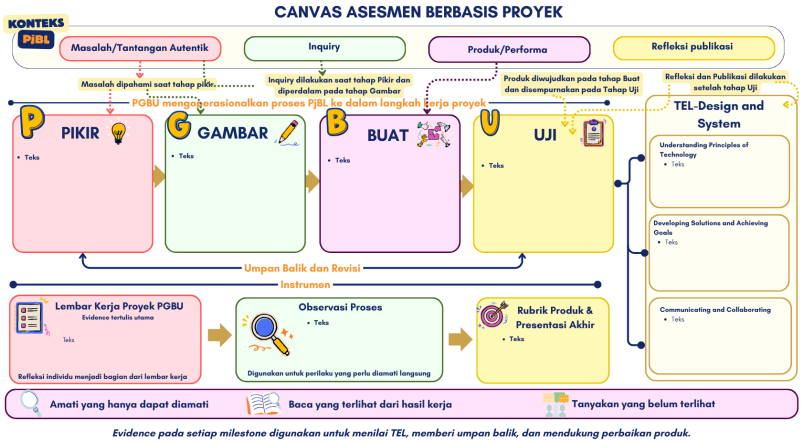
Prosedur Penelitian

Pengumpulan data dilakukan melalui lima tahap. Pertama, guru memperoleh penjelasan mengenai karakteristik PjBL, kedudukan produk sebagai hasil proses inquiry, pentingnya asesmen proses dan produk, tahapan Pikir–Gambar–Buat–Uji (PGBU), dan praktik *Technology and Engineering Literacy* (TEL). Kedua, guru diperkenalkan dengan struktur dan cara menggunakan Canvas Asesmen Berbasis Proyek. Ketiga, setiap guru memilih konteks proyek yang sesuai dengan mata pelajaran atau

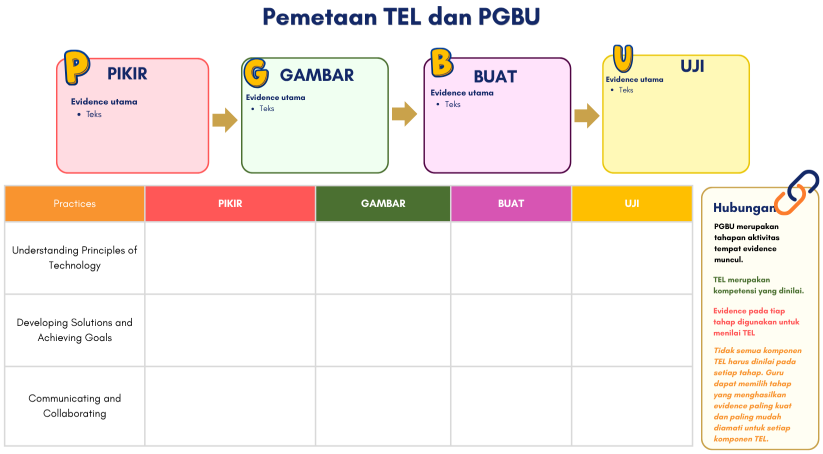
lingkungan pembelajarannya dan menggunakan canvas untuk menyusun rancangan penilaian. Keempat, guru menyerahkan rancangan penilaian pada canvas yang telah diisi. Produk dapat diserahkan dalam format PDF, PowerPoint, atau gambar selama seluruh komponennya dapat dibaca. Kelima, setelah mencoba menggunakan canvas, guru mengisi angket persepsi kebermanfaatan. Produk dan respons angket kemudian dipasangkan menggunakan identitas responden yang telah dinormalisasi. Setiap pasangan data diberi kode anonim G001, G002, dan seterusnya untuk proses analisis.

Canvas Asesmen Berbasis Proyek

Canvas Asesmen Berbasis Proyek merupakan alat bantu perencanaan yang digunakan guru untuk mengorganisasi komponen penilaian dalam satu tampilan visual. Canvas Asesmen Berbasis Proyek yang digunakan dalam penelitian ini terdiri atas dua halaman. Halaman pertama memuat konteks PjBL (masalah atau tantangan autentik, *inquiry*, produk atau performa, serta refleksi dan publikasi), aktivitas pada tahap Pikir, Gambar, Buat, dan Uji beserta alur umpan balik dan revisi, tiga kelompok instrumen (lembar kerja proyek PGBU, observasi proses, serta rubrik produk dan presentasi akhir), dan praktik TEL pada area *Design and Systems*. Halaman kedua memuat bukti utama pada setiap tahap PGBU dan matriks yang memetakan bukti tersebut terhadap tiga praktik TEL Kedua halaman template Canvas Asesmen Berbasis Proyek disajikan pada Gambar 1a dan Gambar 1b.



Gambar 1a. Halaman pertama Template Canvas Asesmen Berbasis Proyek: pemetaan konteks proyek, tujuan pembelajaran, tahapan PGBU, dan bukti belajar.



Gambar 1b. Halaman kedua Template Canvas Asesmen Berbasis Proyek: Bukti Utama Setiap Tahap PGBU dan Pemetaannya terhadap Praktik TEL

Instrumen Penilaian Kualitas Rancangan Penilaian

Kualitas rancangan penilaian guru dinilai menggunakan rubrik analitik yang terdiri atas sembilan dimensi. Setiap dimensi diberi skor 1–4, dengan kriteria 1 = belum tampak, 2 = berkembang, 3 = memadai, dan 4 = sangat kuat.

Tabel 2. Dimensi Rubrik Kualitas Canvas Rancangan Penilaian

Kode	Dimensi yang dinilai	Fokus pemeriksaan
D1	Konteks masalah atau tantangan autentik	Kejelasan masalah, kebutuhan, pengguna, tujuan, dan produk atau performa
D2	Kejelasan dan kesinambungan aktivitas PGBU	Kejelasan tindakan pada tahap Pikir, Gambar, Buat, dan Uji serta kesinambungan antartahap
D3	Kejelasan dan keteramatan bukti belajar	Kekonkretan artefak, data, perilaku, performa, atau refleksi pada setiap tahap
D4	Keselarasan aktivitas PGBU dengan bukti belajar	Kesesuaian bukti sebagai keluaran langsung dari aktivitas pada tahap terkait
D5	Keselarasan bukti dengan metode atau instrumen	Kesesuaian bukti dengan lembar kerja, observasi, rubrik, pertanyaan, dan teknik asesmen lainnya
D6	Ketepatan pemetaan bukti terhadap praktik TEL	Kesesuaian bukti dengan tiga praktik TEL
D7	Keseimbangan cakupan asesmen	Keseimbangan proses dan produk, individu dan kelompok, serta tahap antara dan hasil akhir
D8	Umpan balik, revisi, refleksi, dan komunikasi	Kejelasan titik umpan balik, tindakan revisi, refleksi, presentasi, atau publikasi
D9	Koherensi dan keterlaksanaan rancangan	Konsistensi antarkomponen, kewajaran beban penilaian, kebutuhan sumber daya, dan kemungkinan penerapan

Skor total kualitas rancangan penilaian diperoleh dengan menjumlahkan skor D1–D9 sehingga rentang skor total adalah 9–36. Skor rata-rata dihitung menggunakan persamaan:

$$\bar{X}_{Rancangan\ penilaian} = \frac{\sum_{j=1}^{9} D_j}{9}$$

Rata-rata kualitas rancangan penilaian diinterpretasikan menjadi empat kategori yaitu skor 1,00–1,75 = sangat kurang; lebih dari 1,75–2,50 = kurang; lebih dari 2,50–3,25 = baik; dan lebih dari 3,25–4,00 = sangat baik. Selain skor keseluruhan, skor setiap dimensi dianalisis secara terpisah agar kekuatan dan kelemahan rancangan guru dapat diidentifikasi secara lebih spesifik.

Rubrik disusun berdasarkan kajian literatur dan setiap dimensi dirujuk pada landasan konseptual yang jelas. D1 dan D2 merujuk pada karakteristik PjBL dan alur PGBU (Markula & Aksela, 2022; Suminar et al., 2025); D3–D5 merujuk pada keselarasan tujuan, aktivitas, bukti, dan instrumen penilaian (Harris et al., 2022; Zhao et al., 2023); D6 merujuk pada praktik TEL (National Assessment Governing Board, 2018); D7 dan D8 merujuk pada keseimbangan asesmen proses dan produk serta desain umpan balik (Panadero & Lipnevich, 2022; van der Steen et al., 2023); sedangkan D9 menilai koherensi dan keterlaksanaan rancangan secara keseluruhan.

Setiap dimensi dilengkapi deskriptor untuk masing-masing skor 1–4 sebagai acuan penilaian. Penilaian dilakukan oleh satu penilai terhadap produk yang telah diberi kode anonim. Penilai membaca

kedua halaman canvas secara utuh, memberikan skor pada D1–D9, dan mencatat alasan pemberian skor pada setiap dimensi. Penggunaan satu penilai merupakan keterbatasan penelitian ini dan dibahas pada bagian Keterbatasan Penelitian.

Angket Persepsi Kebermanfaatan

Persepsi guru diukur menggunakan 11 butir pernyataan dengan skala Likert lima kategori, yaitu 1 = sangat tidak setuju, 2 = tidak setuju, 3 = netral, 4 = setuju, dan 5 = sangat setuju. Angket mengukur persepsi guru setelah menggunakan canvas untuk menyusun rancangan penilaian.

Konstruk yang diukur adalah persepsi kebermanfaatan (*perceived usefulness*), yaitu sejauh mana guru meyakini bahwa canvas membantu mereka merancang asesmen berbasis proyek. Konstruk ini berbeda dari persepsi kemudahan penggunaan (*perceived ease of use*), yaitu sejauh mana penggunaan canvas dirasakan tidak memerlukan banyak usaha (Davis, 1989). Sebagian besar butir mengukur kebermanfaatan, sedangkan butir 1 menyangkut kemudahan memahami dan menggunakan canvas. Butir 1 dipertahankan sebagai prasyarat agar canvas dapat dimanfaatkan, tetapi skor total ditafsirkan sebagai persepsi kebermanfaatan. Istilah *usability* tidak digunakan karena angket tidak mengukur efektivitas, efisiensi, dan kepuasan penggunaan secara terpisah.

Butir angket mencakup beberapa aspek, yaitu: (1) kemudahan memahami dan menggunakan canvas; (2) bantuan canvas dalam menyusun asesmen berbasis proyek secara sistematis; (3) penentuan aspek proses dan hasil yang perlu dinilai; (4) keselarasan asesmen dengan tahapan PGBU dan praktik TEL; (5) pengurangan kebingungan dalam menentukan aspek penilaian; (6) efisiensi penyusunan instrumen; (7) penyesuaian penilaian dengan setiap tahap PGBU; (8) keseimbangan penilaian proses dan produk; dan (9) identifikasi bukti belajar yang perlu dikumpulkan.

Sembilan butir pertama dirumuskan secara positif, sedangkan butir 10 dan 11 dirumuskan secara negatif. Butir 10 berbunyi “Canvas asesmen ini tidak membantu saya menentukan aspek yang perlu dinilai”, dan butir 11 berbunyi “Canvas asesmen ini belum membantu saya menilai proses belajar setiap murid secara lebih terarah.” Sebelum analisis, skor kedua butir negatif dibalik menggunakan persamaan:

$$X_R = 6 - X$$

Dengan pembalikan tersebut, skor yang lebih tinggi pada seluruh butir menunjukkan persepsi kebergunaan yang lebih positif.

Analisis Data Kualitas Rancangan Penilaian

Data kualitas rancangan dianalisis menggunakan statistik deskriptif berupa skor total, rata-rata, simpangan baku, nilai minimum–maksimum, kategori kualitas, dan distribusi skor pada setiap dimensi. Analisis per dimensi digunakan untuk mengidentifikasi komponen yang relatif kuat dan komponen yang masih memerlukan pengembangan.

Analisis Rasch terhadap Angket Persepsi

Respons angket dianalisis menggunakan *Rating Scale Model* (RSM) karena seluruh butir menggunakan struktur kategori respons yang sama. Analisis menggunakan estimasi *marginal maximum likelihood*, sedangkan ukuran responden diestimasi dengan metode *expected a posteriori* (EAP). Ukuran butir dan responden dinyatakan dalam satuan logit. Model Rasch memungkinkan skor ordinal dikalibrasi pada skala yang sama sekaligus menilai kecocokan butir, fungsi kategori, reliabilitas, separasi, dan keterarahan butir terhadap responden (Avinç & Doğan, 2024; Carey et al., 2025).

Analisis Rasch dilakukan menggunakan skrip khusus dalam Python versi 3.12.14 dengan dukungan NumPy versi 2.3.5 dan SciPy versi 1.17.0. Estimasi menggunakan kuadratur Gauss–Hermite 61 titik dan optimasi L-BFGS-B.. Visualisasi Peta Wright dibuat menggunakan Matplotlib versi 3.10.8.

Evaluasi model mencakup reliabilitas dan separasi responden, reliabilitas dan separasi butir, nilai infit dan outfit mean square (MNSQ), korelasi butir–ukuran, fungsi kategori respons, keterarahan butir terhadap responden, serta asumsi unidimensionalitas dan independensi lokal. Nilai MNSQ 0,50–1,50 digunakan sebagai rentang kecocokan yang produktif untuk pengukuran. Nilai di atas 1,50 menunjukkan respons yang kurang konsisten dengan model, sedangkan nilai di bawah 0,50 menunjukkan pola yang terlalu mudah diprediksi. Asumsi unidimensionalitas diperiksa melalui analisis komponen utama residual; nilai eigen kontras pertama lebih dari 2,0 ditafsirkan sebagai indikasi

dimensi sekunder. Independensi lokal diperiksa melalui korelasi residual antarbutir; korelasi lebih dari 0,30 ditafsirkan sebagai indikasi ketergantungan lokal. Jumlah strata responden yang dapat dibedakan dihitung dengan rumus H = (4G + 1)/3, dengan G adalah separasi responden (Bond et al., 2020). Keberfungsian kategori ditinjau dari frekuensi penggunaan dan keterurutan ambang kategori; penargetan diperiksa dengan membandingkan lokasi responden dan butir pada skala logit yang sama serta melalui Peta Wright (Noroozi & Karami, 2024; Tan et al., 2024).

Analisis Hubungan Rancangan Penilaian dan Persepsi

Hubungan antara kualitas rancangan penilaian dan persepsi kebermafaatan dianalisis secara eksploratif. Skor rata-rata kualitas rancangan D1–D9 dipasangkan dengan ukuran persepsi Rasch dari guru yang sama dan hubungan kedua variabel dianalisis menggunakan korelasi peringkat Spearman. Koefisien korelasi dilaporkan bersama nilai probabilitas dua arah pada taraf signifikansi 0,05. Interval kepercayaan 95% koefisien korelasi dihitung menggunakan transformasi Fisher *z* untuk menggambarkan presisi estimasi.

HASIL

Kualitas Rancangan Penilaian pada Canvas Asesmen

Analisis dilakukan terhadap 44 produk canvas yang telah memenuhi kriteria kelengkapan, keunikan, keterverifikasian identitas, dan keterpasangan dengan angket persepsi. Produk yang dianalisis mencakup berbagai konteks proyek, antara lain energi terbarukan, teknologi lingkungan, sistem irigasi, alat ukur, keselamatan laboratorium, pertanian, dan perangkat teknologi otomatis. Statistik deskriptif kualitas rancangan disajikan pada Tabel 3.

Tabel 3. Statistik Deskriptif Kualitas Rancangan Penilaian

Indikator	Hasil
Jumlah produk	44
Rata-rata skor total	31,70
Simpangan baku skor total	4,05
Median skor total	33,00
Skor total minimum–maksimum	15–36
Rata-rata skor per dimensi	3,52
Simpangan baku skor per dimensi	0,45
Rata-rata minimum–maksimum	1,67–4,00

Distribusi kategori kualitas rancangan disajikan pada Tabel 4.

Tabel 4. Distribusi Kategori Kualitas Rancangan Penilaian

Rentang rata-rata	Kategori	Frekuensi	Persentase
1,00–1,75	Sangat kurang	1	2,3%
>1,75–2,50	Kurang	0	0,0%

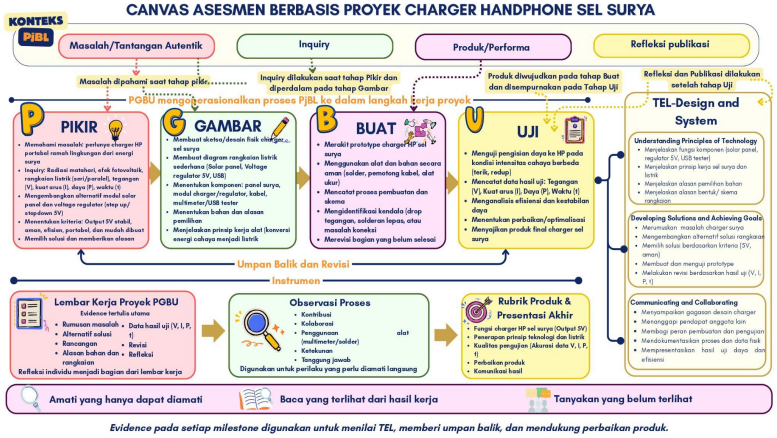
>2,50–3,25	Baik	7	15,9%
>3,25–4,00	Sangat baik	36	81,8%
Total		44	100,0%

Meskipun kualitas rancangan secara keseluruhan cenderung tinggi, terdapat variasi antardimensi. Hasil setiap dimensi disajikan pada Tabel 5.

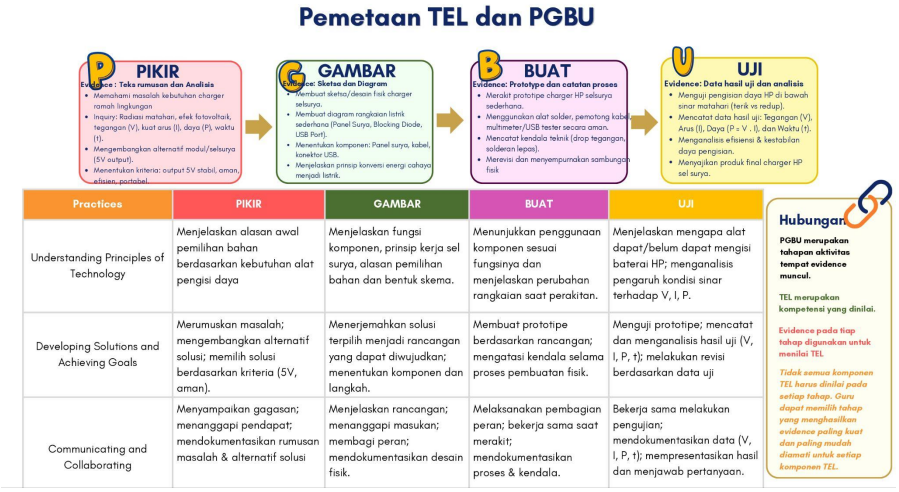
Tabel 5. Kualitas Rancangan Berdasarkan Dimensi Penilaian

Kode	Dimensi yang dinilai	Rata-Rata	SD	Produk dengan Skor 4, n(%)
D1	Konteks masalah atau tantangan autentik	3,39	0,72	22 (50,0%)
D2	Kejelasan dan kesinambungan aktivitas PGBU	3,82	0,45	37 (84,1%)
D3	Kejelasan dan keteramatan bukti belajar	3,82	0,62	40 (90,9%)
D4	Keselarasan aktivitas PGBU dengan bukti belajar	3,86	0,46	40 (90,9%)
D5	Keselarasan bukti dengan metode atau instrumen	3,07	0,55	7 (15,9%)
D6	Ketepatan pemetaan bukti terhadap praktik TEL	3,59	0,69	30 (68,2%)
D7	Keseimbangan cakupan asesmen	3,36	0,57	18 (40,9%)
D8	Umpan balik, revisi, refleksi, dan komunikasi	3,32	0,64	18 (40,9%)
D9	Koherensi dan keterlaksanaan rancangan	3,48	0,70	26 (59,1%)

Untuk melengkapi hasil penilaian kuantitatif, satu produk guru dipilih sebagai ilustrasi penggunaan Canvas Asesmen Berbasis Proyek. Contoh produk disajikan pada Gambar 2a dan Gambar 2b.



Gambar 2a. Halaman pertama contoh Canvas Asesmen Berbasis Proyek yang disusun oleh guru.



Gambar 2b. Halaman kedua contoh Canvas Asesmen Berbasis Proyek yang disusun oleh guru.

Persepsi Guru terhadap Kebermanfaatan Canvas

Sembilan butir positif memperoleh rata-rata antara 4,61–4,77. Butir dengan rata-rata tertinggi adalah butir ketiga, yaitu canvas membantu guru menentukan aspek proses dan hasil yang perlu dinilai (*M=4,77; SD=0,48*). Butir berikutnya adalah mengenai bantuan canvas dalam merancang asesmen berbasis proyek secara lebih sistematis (*M=4,75;SD=0,44*) dan butir kelima yaitu mengenai pengurangan kebingungan dalam menentukan aspek penilaian (*M=4,70; SD=0,46*).

Rata-rata terendah di antara butir positif ditemukan pada butir 1 mengenai kemudahan memahami dan menggunakan canvas serta butir 8 mengenai keseimbangan penilaian proses dan hasil. Kedua butir tersebut memperoleh rata-rata 4,61. Meskipun menjadi nilai terendah di antara butir positif, nilainya tetap menunjukkan tingkat persetujuan yang tinggi.

Setelah skor dibalik, dua butir negatif menghasilkan rata-rata yang lebih rendah dan variasi respons yang lebih besar, yaitu butir 10 (*M=3,32; SD=1,78*), sedangkan butir 11 (*M=3,18; SD=1,74*). Pola respons pada kedua butir tersebut lebih tersebar dibandingkan dengan sembilan butir positif.

Hasil Analisis Rasch

Analisis Rasch dilakukan terhadap 11 butir angket. Secara deskriptif, persepsi guru terhadap penggunaan Canvas Asesmen Berbasis Proyek cenderung positif. Dari 484 respons butir, sebanyak 310 respons (64,0%) berada pada kategori sangat setuju dan 127 respons (26,2%) pada kategori setuju. Dengan demikian, 90,3% respons terkonsentrasi pada dua kategori persetujuan tertinggi. Hasil analisis *Rating Scale Model* disajikan pada Tabel 6.

Tabel 6. Ringkasan Hasil Analisis Rasch pada 11 Butir Persepsi

Indikator	Hasil	Interpretasi
Jumlah responden	44	Sampel akhir berpasangan
Jumlah butir	11	Sembilan butir positif dan dua butir negatif yang dibalik
Rata-rata ukuran responden	1,857 logit	Persepsi guru berada di atas rata-rata lokasi butir
Reliabilitas responden	0,816	Memadai untuk menggambarkan perbedaan pada tingkat kelompok
Separasi responden	2,105	Angket cukup mampu membedakan beberapa tingkat persepsi guru
Reliabilitas butir	0,829	Urutan relatif lokasi butir cukup stabil pada sampel
Separasi butir	2,198	Terdapat perbedaan lokasi relatif antarputir
Cronbach's alpha	0,755	Konsistensi internal memadai
Responden dengan skor maksimum	11 guru (25,0%)	Menunjukkan efek plafon
Kontras residual pertama	4,323	Mengindikasikan adanya pola sekunder dalam respons
Korelasi residual tertinggi	I6–I8 = 0,819	Menunjukkan ketergantungan lokal antarputir
Korelasi residual butir negatif	I10_R–I11_R = 0,810	Menunjukkan kemungkinan efek kesamaan redaksi negatif
Keterurutan ambang kategori	Tidak berurutan	Lima kategori respons belum berfungsi secara optimal

Reliabilitas responden 0,816 dan separasi 2,105 menunjukkan bahwa angket dapat membedakan sekitar tiga strata persepsi guru (H = 3,14). Namun, beberapa indikator menunjukkan bahwa kualitas pengukuran belum optimal. Nilai eigen kontras residual pertama sebesar 4,323 melampaui batas 2,0 sehingga terdapat indikasi dimensi sekunder. Korelasi residual I6–I8 (0,819) dan I10_R–I11_R (0,810) jauh melampaui 0,30 sehingga menunjukkan ketergantungan lokal. Ambang kategori yang tidak berurutan juga menunjukkan bahwa lima kategori respons belum digunakan secara berjenjang sebagaimana diharapkan. Statistik setiap butir disajikan pada Tabel 7 dan fungsi kategori respons disajikan pada Tabel 8.

Tabel 7. Statistik Butir pada Model Rasch 11 Butir

Butir	Ukuran (logit)	SE	*Infit* MNSQ	*Outfit* MNSQ	Korelasi *point–measure*	Interpretasi
I11_R	1,634	0,214	1,333	1,205	0,701	Produktif
I10_R	1,511	0,214	1,483	1,175	0,675	Produktif
I1	−0,148	0,323	0,553	0,656	0,546	Produktif
I8	−0,148	0,323	0,483	0,462	0,637	*Overfit*
I4	−0,215	0,333	0,470	0,549	0,597	*Overfit*
I6	−0,215	0,333	0,695	0,666	0,561	Produktif
I9	−0,287	0,343	0,494	0,495	0,603	*Overfit*
I7	−0,287	0,343	0,573	0,558	0,586	Produktif
I5	−0,449	0,366	0,472	0,392	0,658	*Overfit*
I2	−0,642	0,394	0,599	0,476	0,572	*Overfit*
I3	−0,754	0,411	0,787	0,478	0,565	*Overfit*

Tabel 8. Fungsi Kategori Respons pada Model Rasch 11 Butir

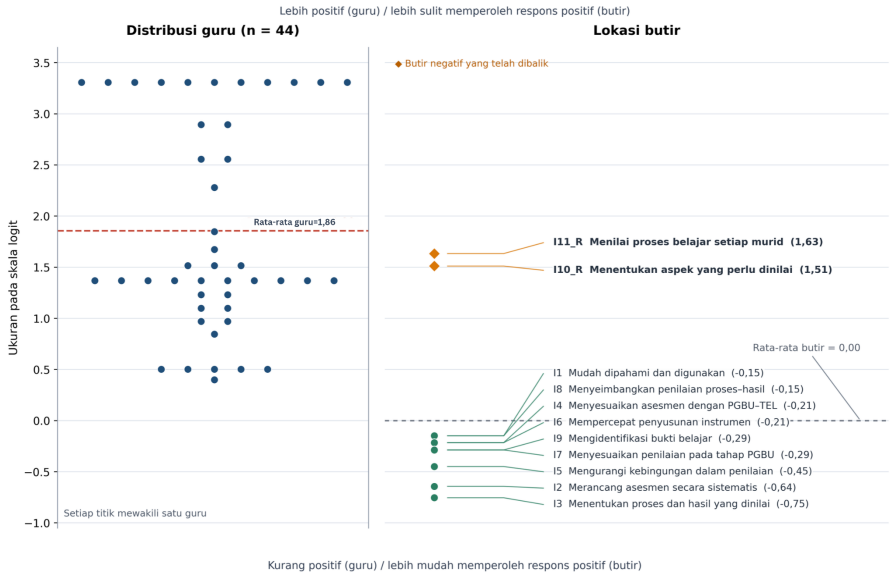
Kategori	Frekuensi	Persentase (%)	Rata-rata ukuran teramati (logit)	Langkah kategori	Ambang kategori (logit)
1	24	5,0	1,291	1 → 2	0,087
2	15	3,1	0,802	2 → 3	0,827
3	8	1,7	0,731	3 → 4	−1,836
4	127	26,2	1,035	4 → 5	0,922
5	310	64,0	2,317	–	–
Total	**484**	**100,0**	–	–	–

Tabel 7 menunjukkan bahwa ukuran sembilan butir positif berada pada rentang −0,754 sampai −0,148 logit, sedangkan dua butir negatif yang dibalik berada jauh lebih tinggi (1,511 dan 1,634 logit). Butir yang paling mudah disetujui adalah I3, yaitu canvas membantu menentukan aspek proses dan hasil yang perlu dinilai, sedangkan butir yang paling sulit disetujui adalah I11_R yang berkaitan dengan penilaian proses belajar setiap peserta didik. Tidak ada butir dengan nilai infit atau outfit MNSQ di atas 1,50, sehingga tidak ditemukan butir yang mengganggu pengukuran. Lima butir (I1, I6, I7, I10_R, dan I11_R) berada pada rentang produktif, sedangkan enam butir (I2, I3, I4, I5, I8, dan I9) memiliki nilai infit atau outfit di bawah 0,50 (*overfit*). *Overfit* menunjukkan pola respons yang terlalu mudah diprediksi, sehingga butir kurang memberikan informasi baru, tetapi tidak merusak pengukuran (Bond et al., 2020). Pola ini sejalan dengan Cronbach's alpha sembilan butir positif yang sangat tinggi (0,959) dan efek plafon, yang mengindikasikan kemiripan isi antarbutir positif. Seluruh korelasi butir–ukuran bernilai positif (0,546–0,701), sehingga seluruh butir, termasuk dua butir negatif setelah dibalik, searah dengan konstruk yang diukur.

Tabel 8 menunjukkan bahwa respons terkonsentrasi pada kategori 4 dan 5 (90,3%), sedangkan kategori 1–3 hanya digunakan pada 47 respons (9,7%). Kategori 3 hanya muncul delapan kali, di bawah batas minimum 10 observasi per kategori. Rata-rata ukuran teramati juga tidak meningkat secara berurutan: respons pada kategori 1 berasal dari guru dengan rata-rata ukuran 1,291 logit, lebih tinggi daripada kategori 2 (0,802), 3 (0,731), dan 4 (1,035). Karena sembilan butir positif tidak memperoleh respons pada kategori 1 dan 2, seluruh respons pada kedua kategori tersebut berasal dari butir negatif setelah dibalik. Artinya, sebagian guru yang secara umum berpersepsi positif tetap menyetujui pernyataan negatif, yang mengindikasikan kekeliruan membaca redaksi negatif atau kecenderungan menyetujui pernyataan. Ambang kategori juga tidak berurutan karena ambang 3→4 (−1,836 logit) berada di bawah ambang 1→2 (0,087) dan 2→3 (0,827). Setelah kedua butir negatif dikeluarkan,

ambang kategori menjadi berurutan (Tabel 9), sehingga ketidakberurutan ini terutama bersumber dari butir negatif. Temuan ini memperkuat perlunya meninjau redaksi butir negatif dan jumlah kategori respons pada pengembangan angket berikutnya.

Untuk memperlihatkan kesesuaian antara tingkat persepsi guru dan lokasi butir angket, hasil analisis Rasch divisualisasikan menggunakan Peta Wright pada Gambar 3.



Gambar 3. Peta Wright Persepsi Guru dan Lokasi Butir Angket Kebermanfaatan Canvas Asesmen Berbasis Proyek

Gambar 3 menunjukkan bahwa distribusi ukuran persepsi guru cenderung berada di atas rata-rata lokasi butir. Distribusi responden yang lebih tinggi daripada sebagian besar lokasi butir menunjukkan bahwa guru secara umum memiliki persepsi positif terhadap penggunaan canvas. Namun, terbatasnya butir pada tingkat logit tinggi menunjukkan bahwa angket belum cukup tertarget untuk membedakan guru dengan persepsi yang sangat positif. Pola ini konsisten dengan ditemukannya 11 guru atau 25,0% responden yang memperoleh skor maksimum.

Analisis Sensitivitas terhadap Butir Negatif

Untuk memeriksa pengaruh redaksi negatif, dilakukan analisis sensitivitas dengan mengeluarkan I10 dan I11. Perbandingan model 11 butir dengan model sembilan butir positif disajikan pada Tabel 9.

Tabel 9. Perbandingan Model Utama dan Analisis Sensitivitas

Indikator	Model 11 Butir	Model 9 Butir
Reliabilitas responden	0,816	0,815
Separasi responden	2,105	2,099
Cronbach’s alpha	0,755	0,959
Rata-rata responden–butir	1,857 logit	3,110 logit

Kontras residual pertama	4,323	1,172
Korelasi residual tertinggi	0,819	0,696
Responden dengan skor maksimum	11 (25,0%)	23 (52,3%)
Kategori yang teramati	1–5	3–5
Ambang kategori berurutan	Tidak	Ya

Setelah kedua butir negatif dikeluarkan, kontras residual pertama menurun dari 4,323 menjadi 1,172. Penurunan tersebut menunjukkan bahwa kesesuaian sembilan butir positif terhadap satu konstruk menjadi lebih baik. Cronbach's alpha juga meningkat menjadi 0,959, yang menunjukkan konsistensi internal sangat tinggi. Namun, nilai yang sangat tinggi ini juga dapat menunjukkan bahwa beberapa butir positif memiliki isi atau pola respons yang terlalu serupa.

Penghapusan butir negatif juga memperkuat efek plafon. Jumlah guru dengan skor maksimum meningkat dari 11 orang (25,0%) menjadi 23 orang (52,3%), sedangkan selisih rata-rata ukuran responden dan butir meningkat menjadi 3,110 logit. Selain itu, pada sembilan butir positif tidak ditemukan respons pada kategori 1 dan 2. Dengan demikian, model sembilan butir memang lebih homogen, tetapi semakin kurang mampu membedakan guru dengan tingkat persepsi yang sama-sama tinggi.

Secara keseluruhan, hasil Rasch mendukung temuan bahwa guru memiliki persepsi positif terhadap penggunaan Canvas Asesmen Berbasis Proyek. Guru terutama merasakan bahwa canvas membantu menentukan aspek proses dan hasil yang perlu dinilai serta membantu menyusun asesmen secara lebih sistematis.

Hubungan antara Persepsi Kebermanfaatan dan Kualitas Rancangan

Hasil uji korelasi Spearman ditunjukan pada Tabel 10.

Tabel 10. Hubungan antara Persepsi Kebermanfaatan dan Kualitas Rancangan

Variabel yang dikorelasikan	n	Koefisien Spearman (r_s)	*p*	Interpretasi
Ukuran persepsi Rasch dan skor kualitas rancangan	44	0,087	0,576	Hubungan positif sangat lemah dan tidak signifikan

Koefisien korelasi bernilai positif, tetapi sangat lemah dan tidak signifikan (r_s = 0,087; p = 0,576; IK 95% [−0,22; 0,37]). Interval kepercayaan yang lebar menunjukkan bahwa data belum cukup presisi untuk menyimpulkan tidak adanya hubungan sama sekali; yang dapat disimpulkan adalah tidak ditemukannya hubungan yang kuat antara persepsi kebermanfaatan dan kualitas rancangan.

PEMBAHASAN

Canvas sebagai Penyangga Perencanaan Asesmen Berbasis Proyek

Kualitas rancangan guru secara keseluruhan berada pada kategori sangat baik, dengan rata-rata 3,52 dari skala 4. Sebanyak 36 produk (81,8%) termasuk kategori sangat baik dan tujuh produk (15,9%) termasuk kategori baik. Dimensi yang memperoleh skor tertinggi adalah keselarasan antara aktivitas PGBU dan bukti belajar (D4), diikuti kejelasan aktivitas PGBU (D2) serta kejelasan dan keteramatan bukti belajar (D3). Pola ini menunjukkan bahwa guru relatif berhasil menjabarkan kegiatan peserta didik pada tahap Pikir, Gambar, Buat, dan Uji serta menentukan bukti yang dapat dihasilkan pada setiap

tahap. Artinya, canvas membantu menggeser perhatian penilaian dari produk akhir menuju proses belajar yang berlangsung selama proyek.

Temuan ini penting karena PjBL tidak hanya ditandai oleh kegiatan menghasilkan produk. PjBL juga melibatkan penyelidikan masalah autentik, kolaborasi, penerapan pengetahuan, pengembangan artefak, penyajian hasil, dan refleksi. Karena itu, penilaian perlu menangkap proses yang mengantarkan peserta didik pada produk, bukan hanya menilai kualitas produk akhir. Karakter tersebut konsisten dengan temuan Markula dan Aksela (2022) serta Pupik Dean et al. (2023) mengenai pentingnya masalah, inquiry, kolaborasi, artefak, presentasi, dan refleksi dalam pelaksanaan PjBL. Meta-analisis Zhang dan Ma (2023) juga menunjukkan bahwa hasil PjBL bersifat multidimensional sehingga memerlukan bukti dan metode penilaian yang beragam.

Tingginya skor ketepatan pemetaan bukti terhadap praktik TEL (D6; rata-rata 3,59) juga menunjukkan bahwa sebagian besar guru dapat menghubungkan bukti belajar dengan praktik memahami prinsip teknologi, mengembangkan solusi dan mencapai tujuan, serta berkomunikasi dan berkolaborasi. Struktur tersebut memungkinkan praktik TEL yang bersifat abstrak diterjemahkan menjadi bukti yang lebih teramati. Meskipun demikian, hasil ini belum membuktikan bahwa pemetaan tersebut akan terlaksana secara konsisten di kelas. Produk yang dianalisis masih berupa rancangan, sehingga kualitas implementasi dan bukti belajar peserta didik belum dapat dipastikan.

Kesenjangan antara Penentuan Bukti dan Penyusunan Instrumen

Meskipun kualitas rancangan secara umum tinggi, dimensi keselarasan antara bukti dengan metode atau instrumen penilaian (D5) memperoleh skor paling rendah, yaitu 3,07. Skor ini masih berada pada kategori baik, tetapi lebih rendah dibandingkan kejelasan aktivitas, bukti belajar, dan keselarasan aktivitas dengan bukti. Hasil tersebut menunjukkan bahwa guru relatif lebih mudah menentukan apa yang dilakukan peserta didik dan bukti apa yang dihasilkan daripada menetapkan bagaimana bukti tersebut dikumpulkan dan dinilai.

Perbedaan tersebut menggambarkan tantangan dalam literasi asesmen guru. Kompetensi asesmen tidak berhenti pada kemampuan mengenali tujuan dan bukti belajar; guru juga perlu memilih metode, menentukan kriteria, menafsirkan hasil, dan merancang tindak lanjut. Kajian Pastore (2023) dan Lei dan Lei (2026) menunjukkan bahwa penerjemahan pengetahuan asesmen ke praktik kontekstual masih menjadi persoalan. Model Yan dan Pastore (2022) juga menegaskan keterkaitan dimensi konseptual, praktis, dan sosioemosional. Karena itu, alat bantu visual tetap memerlukan pendampingan profesional agar rancangan menjadi terarah dan operasional (Baartman & Gulikers, 2025).

Dalam beberapa produk, guru telah mencantumkan lembar kerja, observasi, rubrik, presentasi, refleksi, atau penilaian diri. Namun, pencantuman nama instrumen belum selalu diikuti penjelasan mengenai aspek yang diperiksa, kriteria keberhasilan, waktu penggunaannya, dan hubungan langsungnya dengan bukti tertentu. Sebagai contoh, bukti berupa sketsa rancangan memerlukan kriteria yang berbeda dari bukti berupa proses kolaborasi, data hasil pengujian, atau produk akhir. Apabila semua bukti hanya diarahkan pada satu rubrik produk, proses berpikir, pengambilan keputusan, dan perbaikan solusi berisiko tidak terukur secara memadai.

Rendahnya skor D5 dapat dijelaskan oleh beberapa kemungkinan. Pertama, penyusunan instrumen merupakan langkah yang paling menuntut dalam perencanaan asesmen karena guru harus menerjemahkan bukti menjadi indikator, kriteria, dan tingkat capaian yang dapat diskor; kajian literasi asesmen menunjukkan bahwa langkah penerjemahan ini lebih sulit daripada mengenali tujuan dan bukti belajar (Yan & Pastore, 2022; Pastore, 2023). Kedua, struktur canvas turut membentuk pola tersebut. Bagian instrumen pada canvas berupa tiga kelompok instrumen, yaitu lembar kerja proyek, observasi proses, serta rubrik produk dan presentasi akhir, tetapi belum menyediakan ruang untuk menuliskan indikator, kriteria keberhasilan, dan waktu penggunaan instrumen bagi setiap bukti. Akibatnya, guru cenderung hanya menuliskan nama instrumen. Ketiga, rancangan disusun dalam waktu kegiatan pengembangan profesional yang terbatas, sedangkan pengembangan instrumen yang operasional umumnya memerlukan proses tersendiri. Keempat, kebiasaan menilai proyek dengan satu rubrik produk dapat mendorong guru menggabungkan beragam bukti ke dalam satu instrumen. Temuan van der Steen et al. (2023) bahwa ketidakselarasan antara bukti dan tindak lanjut masih muncul pada rancangan guru berpengalaman menunjukkan bahwa kesulitan ini tidak semata-mata disebabkan oleh kurangnya pengalaman. Karena penelitian ini tidak mengumpulkan data proses berpikir guru, penjelasan tersebut masih perlu dikonfirmasi melalui wawancara atau analisis protokol. Implikasinya, versi canvas

berikutnya perlu menambahkan kolom indikator atau kriteria serta waktu pengumpulan bukti untuk setiap instrumen.

Dimensi keseimbangan cakupan asesmen (D7; rata-rata 3,36) dan umpan balik, revisi, refleksi, serta komunikasi (D8; rata-rata 3,32) juga relatif lebih rendah daripada D2–D4. Hal ini menunjukkan bahwa sebagian guru masih perlu memperjelas keseimbangan antara penilaian proses dan produk, individu dan kelompok, serta formatif dan sumatif. Pada proyek kelompok, kualitas rancangan tidak selalu menggambarkan kontribusi dan proses belajar setiap peserta didik. Oleh karena itu, guru memerlukan bukti individual, seperti catatan keputusan, jurnal refleksi, penjelasan lisan, hasil observasi, atau tanggapan terhadap umpan balik.

Asesmen formatif tidak cukup hanya dengan mencantumkan bahwa guru akan memberikan umpan balik. Rancangan perlu menentukan kapan umpan balik diberikan, bukti apa yang diperiksa, siapa yang memberi umpan balik, dan tindakan perbaikan apa yang dilakukan peserta didik. Umpan balik berfungsi untuk pembelajaran apabila peserta didik dapat memahami, menggunakan, dan menindaklanjutinya sebelum pekerjaan selesai. Karena itu, desain umpan balik perlu menghubungkan pesan, konteks, pelaksana, dan tindakan perbaikan secara eksplisit (Panadero & Lipnevich, 2022; de Kleijn, 2023; Woitt et al., 2025).

Persepsi Guru terhadap Kebermanfaatan Canvas

Hasil angket menunjukkan bahwa guru memiliki persepsi yang positif terhadap penggunaan canvas. Sebanyak 90,3% respons berada pada kategori setuju dan sangat setuju. Peta Wright juga menunjukkan bahwa rata-rata ukuran persepsi guru sebesar 1,857 logit berada di atas rata-rata lokasi butir sebesar 0 logit. Temuan ini menunjukkan bahwa sebagian besar pernyataan relatif mudah memperoleh respons positif.

Butir 3 merupakan butir yang paling mudah memperoleh persetujuan tinggi. Dengan demikian, manfaat yang paling kuat dirasakan guru adalah bantuan canvas dalam menentukan aspek proses dan hasil yang perlu dinilai. Butir 2 menjadi butir berikutnya yang paling mudah disetujui, yaitu bantuan canvas dalam merancang asesmen berbasis proyek secara lebih sistematis. Temuan persepsi tersebut sejalan dengan hasil analisis produk, khususnya tingginya skor kejelasan bukti belajar dan keselarasan aktivitas PGBU dengan bukti. Keselarasan antara dua sumber data ini memberikan indikasi bahwa struktur visual canvas membantu guru mengorganisasi komponen perencanaan yang sebelumnya mungkin tersebar dalam beberapa dokumen.

Butir 10 dan butir 11 memiliki pola yang berbeda dari sembilan butir positif. Kedua butir tersebut relatif lebih sulit memperoleh respons positif dan memiliki korelasi residual yang tinggi. Setelah keduanya dikeluarkan, struktur residual menjadi lebih baik, tetapi efek plafon meningkat. Temuan ini menunjukkan kemungkinan adanya efek redaksi negatif. Indikasi ini diperkuat oleh perbandingan butir 3 dan butir 10 yang isinya hampir sama, yaitu bantuan canvas dalam menentukan aspek yang perlu dinilai. Dalam redaksi positif, butir 3 merupakan butir yang paling mudah disetujui (−0,75 logit), sedangkan dalam redaksi negatif, butir 10 termasuk butir yang paling sulit memperoleh respons positif (1,51 logit). Perbedaan lokasi sebesar 2,26 logit pada isi yang serupa menunjukkan bahwa respons guru dipengaruhi oleh arah redaksi, bukan semata-mata oleh isi butir. Penelitian Dueber et al. (2022) memperlihatkan bahwa pencampuran butir positif dan negatif dapat memengaruhi struktur dimensi, ambang respons, rata-rata skor, dan reliabilitas instrumen.

Butir 11 yang paling sulit disetujui berkaitan dengan penilaian proses belajar setiap peserta didik. Pola ini sejalan dengan skor D7 yang relatif rendah pada analisis produk, yaitu keseimbangan antara penilaian individu dan kelompok. Meskipun demikian, keterkaitan ini perlu ditafsirkan secara hati-hati karena lokasi butir 11 juga dipengaruhi oleh redaksi negatifnya

Hubungan antara Persepsi Kebermanfaatan dan Kualitas Rancangan

Korelasi antara ukuran persepsi Rasch dan kualitas rancangan sangat lemah serta tidak signifikan, ($r_s = 0{,}087;\ p = 0{,}576$). Artinya, guru yang memberikan persepsi lebih positif terhadap canvas tidak selalu menghasilkan rancangan penilaian dengan skor lebih tinggi. Sebaliknya, guru yang menghasilkan produk berkualitas tinggi juga tidak selalu memberikan skor persepsi paling tinggi. Tidak ditemukannya hubungan juga sejalan dengan penelitian mengenai keterbatasan penilaian diri sebagai pengganti ukuran kinerja. Ernst et al. (2023) menunjukkan bahwa hubungan antara penilaian diri calon guru dan ukuran pengetahuan objektif dapat lemah atau bahkan tidak ditemukan, serta bergantung pada

domain dan kondisi pengukurannya. Karena itu, persepsi positif tidak dapat digunakan sebagai bukti tunggal bahwa seseorang telah menunjukkan kompetensi yang setara dalam produk atau praktiknya.

Beberapa faktor dapat menjelaskan tidak signifikannya hubungan tersebut. Pertama, kedua variabel memiliki rentang yang terbatas: 81,8% rancangan berada pada kategori sangat baik dan 25,0% guru memperoleh skor persepsi maksimum. Efek plafon pada kedua variabel mengurangi variasi data sehingga koefisien korelasi cenderung melemah. Kedua, kedua variabel mengukur konstruk yang berbeda. Persepsi kebermanfaatan menggambarkan penilaian guru terhadap bantuan yang diberikan alat, sedangkan kualitas rancangan menggambarkan kemampuan guru menghasilkan rancangan yang selaras. Guru dapat menilai canvas sangat bermanfaat justru karena menyadari keterbatasannya dalam merancang asesmen, sementara guru yang sudah terampil dapat menghasilkan rancangan yang baik tanpa merasa terlalu terbantu. Ketiga, keterbatasan pengukuran turut berperan. Angket menunjukkan indikasi dimensi sekunder, ketergantungan lokal, dan ambang kategori yang tidak berurutan, sedangkan kualitas rancangan dinilai oleh satu penilai; kesalahan pengukuran pada kedua variabel dapat menurunkan koefisien korelasi. Keempat, dengan 44 responden, daya uji untuk mendeteksi korelasi sedang (r = 0,30) hanya sekitar 50%, dan interval kepercayaan 95% koefisien yang diperoleh (−0,22 sampai 0,37) masih mencakup hubungan lemah hingga sedang. Oleh karena itu, hasil ini lebih tepat ditafsirkan sebagai tidak ditemukannya bukti hubungan, bukan sebagai bukti bahwa persepsi dan kualitas rancangan sama sekali tidak berhubungan.

Implikasi Pengembangan dan Penggunaan Canvas

Secara praktis, Canvas Asesmen Berbasis Proyek dapat digunakan sebagai alat bantu awal untuk mengarahkan guru melihat asesmen sebagai rangkaian pengumpulan bukti sepanjang proyek. Struktur PGBU memberikan jalur yang cukup konkret bagi guru untuk memetakan aktivitas dan bukti belajar, sedangkan praktik TEL membantu menunjukkan kemampuan teknologi dan rekayasa yang dapat muncul melalui proyek. Canvas berfokus pada hubungan aktivitas, bukti, metode penilaian, dan praktik TEL, maka penggunaannya dapat disesuaikan dengan berbagai mata pelajaran dan konteks proyek. Namun, keterterapan lintas mata pelajaran dalam penelitian ini baru terlihat pada variasi rancangan guru dan belum dibuktikan melalui implementasi. Setiap mata pelajaran memerlukan penyesuaian konsep, bentuk bukti, kriteria kualitas, dan praktik khas disiplin. Literatur mengenai asesmen berbasis perkembangan belajar juga menekankan perlunya menyeimbangkan kerangka yang terstruktur dengan penyesuaian konteks dan kebutuhan peserta didik (Harris et al., 2022; Baartman & Gulikers, 2025).

Keterbatasan Penelitian

Penelitian ini secara khusus menilai kualitas rancangan penilaian yang disusun guru menggunakan canvas. Sesuai dengan ruang lingkup tersebut, penelitian ini tidak mengevaluasi keterlaksanaan rancangan di kelas maupun pengaruhnya terhadap pembelajaran peserta didik. Penelitian selanjutnya dapat memperluas kajian melalui pengujian implementasi canvas, observasi pembelajaran, analisis bukti belajar peserta didik, serta dokumentasi proses umpan balik dan revisi.

Selain itu, kualitas rancangan dinilai oleh satu penilai sehingga reliabilitas antarpenilai belum dapat diestimasi. Angket persepsi juga masih menunjukkan efek plafon, indikasi dimensi sekunder, dan ambang kategori yang tidak berurutan, sehingga perlu direvisi dengan menambahkan butir yang lebih sulit disetujui, menyusun ulang butir berredaksi negatif, dan meninjau jumlah kategori respons. Sampel berasal dari satu kegiatan pengembangan profesional dan ditetapkan berdasarkan kriteria kelengkapan data, sehingga temuan belum dapat digeneralisasikan secara luas.

KESIMPULAN

Penelitian ini menghasilkan tiga temuan utama pada tahap perancangan. Pertama, guru dapat menggunakan Canvas Asesmen Berbasis Proyek untuk menyusun rancangan penilaian dengan rerata kualitas 3,52 dari skala 4; 81,8% rancangan berada pada kategori sangat baik. Kekuatan rancangan terletak pada keselarasan aktivitas PGBU dengan bukti belajar (3,86), kejelasan aktivitas PGBU (3,82), dan kejelasan bukti belajar (3,82), sedangkan kelemahannya terletak pada keselarasan bukti dengan metode atau instrumen penilaian (3,07), umpan balik dan refleksi (3,32), serta keseimbangan cakupan asesmen (3,36). Kedua, guru mempersepsikan canvas bermanfaat, terutama untuk menentukan aspek proses dan hasil yang perlu dinilai serta untuk menyusun asesmen secara sistematis; 90,3% respons berada pada kategori setuju dan sangat setuju, meskipun angket masih menunjukkan efek plafon.

Ketiga, persepsi kebermanfaatan tidak berkorelasi signifikan dengan kualitas rancangan (r_s = 0,087; p = 0,576), sehingga persepsi positif tidak dapat dijadikan bukti tunggal kualitas rancangan.

Kontribusi penelitian ini terletak pada dua hal. Secara praktis, penelitian ini menawarkan alat bantu perencanaan yang menautkan tahapan PGBU, bukti proses dan produk, instrumen penilaian, serta praktik TEL dalam satu tampilan, disertai bukti awal mengenai bagian rancangan yang sudah dikuasai guru dan bagian yang masih memerlukan pendampingan. Secara metodologis, penelitian ini menunjukkan pentingnya mengevaluasi alat bantu perencanaan melalui kualitas produk yang dihasilkan, tidak hanya melalui persepsi pengguna. Temuan ini terbatas pada kualitas rancangan dan tidak dapat digunakan untuk menyimpulkan efektivitas canvas terhadap proses maupun hasil belajar peserta didik.

Oleh karena itu, penggunaan canvas perlu disertai pendampingan lanjutan, contoh pengisian yang memadai, penambahan kolom indikator atau kriteria pada bagian instrumen, serta proses penelaahan dan pemberian umpan balik terhadap rancangan guru. Pendampingan tersebut diperlukan agar hasil pemetaan dalam canvas dapat dikembangkan menjadi perangkat penilaian yang operasional dan selaras dengan tujuan pembelajaran. Karena penelitian ini masih terbatas pada tahap perancangan, penelitian selanjutnya perlu mengimplementasikan canvas dalam pembelajaran di kelas untuk menguji kepraktisan, konsistensi penerapan, serta kontribusinya terhadap kualitas asesmen proses dan produk dalam pembelajaran berbasis proyek.

REFERENSI

Abd Halim, H., Hamzah, M. I., & Zulkifli, H. (2024). A systematic review on the formative assessment practice in teaching and learning in secondary school. *International Journal of Evaluation and Research in Education, 13*(2), 1173–1183. https://doi.org/10.11591/ijere.v13i2.26187

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the Rasch model. *Education and Information Technologies, 29*, 22895–22941. https://doi.org/10.1007/s10639-024-12662-7

Baartman, L. K. J., & Gulikers, J. T. M. (2025). VET teacher professional development in formative assessment using learning progressions. *Journal of Formative Design in Learning, 9*, 136–148. https://doi.org/10.1007/s41686-025-00107-4

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the Assessment Design Decisions Framework. *Teaching in Higher Education, 21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. https://doi.org/10.4324/9780429030499

Børte, K., & Lillejord, S. (2024). Learning to teach: Aligning pedagogy and technology in a learning design tool. *Teaching and Teacher Education, 148*, Article 104693. https://doi.org/10.1016/j.tate.2024.104693

Carey, M. D., Martin, D. A., & McMaster, N. (2025). The development and validation of a self-audit survey instrument that evaluates preservice teachers' confidence to use technologies to support student learning. *International Journal of Research & Method in Education, 48*(2), 160–182. https://doi.org/10.1080/1743727X.2024.2341714

Crespí, P., García-Ramos, J. M., & Queiruga-Dios, M. (2022). Project-based learning (PBL) and its impact on the development of interpersonal competences in higher education. *Journal of New Approaches in Educational Research, 11*(2), 259–276. https://doi.org/10.7821/naer.2022.7.993

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

de Kleijn, R. A. M. (2023). Supporting student and teacher feedback literacy: An instructional model for student feedback processes. *Assessment & Evaluation in Higher Education, 48*(2), 186–200. https://doi.org/10.1080/02602938.2021.1967283

Dueber, D. M., Toland, M. D., Lingat, J. E., Love, A. M. A., Qiu, C., Wu, R., & Brown, A. V. (2022). To reverse item orientation or not to reverse item orientation, that is the question. *Assessment, 29*(7), 1422–1440. https://doi.org/10.1177/10731911211017635

Ernst, H. M., Wittwer, J., & Voss, T. (2023). Do they know what they know? Accuracy in teacher candidates' self-assessments and its influencing factors. *British Educational Research Journal, 49*(4), 649–673. https://doi.org/10.1002/berj.3860

Halawa, S., Lin, T.-C., & Hsu, Y.-S. (2024). Exploring instructional design in K–12 STEM education: A systematic literature review. *International Journal of STEM Education, 11*, Article 43. https://doi.org/10.1186/s40594-024-00503-5

Harris, L. R., Adie, L., & Wyatt-Smith, C. (2022). Learning progression–based assessments: A systematic review of student and teacher uses. *Review of Educational Research, 92*(6), 996–1040. https://doi.org/10.3102/00346543221081552

Kidd, N., Tang, K.-S., & Lin, K.-Y. (2026). A systematic review of engineering practice in K–12 STEM studies. *Research in Science & Technological Education, 44*(3), 1043–1067. https://doi.org/10.1080/02635143.2025.2561994

Lavado-Anguera, S., Velasco-Quintana, P.-J., & Terrón-López, M.-J. (2024). Project-based learning (PBL) as an experiential pedagogical methodology in engineering education: A review of the literature. *Education Sciences, 14*(6), Article 617. https://doi.org/10.3390/educsci14060617

Lei, W., & Lei, Z. (2026). Formative assessment literacy: A systematic review. *Language Testing in Asia, 16*, Article 7. https://doi.org/10.1186/s40468-025-00418-0

Markula, A., & Aksela, M. (2022). The key characteristics of project-based learning: How teachers implement projects in K–12 science education. *Disciplinary and Interdisciplinary Science Education Research, 4*, Article 2. https://doi.org/10.1186/s43031-021-00042-x

National Assessment Governing Board. (2018). *Technology and engineering literacy framework for the 2018 National Assessment of Educational Progress*. U.S. Department of Education. https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/technology/2018-technology-framework.pdf

Noroozi, S., & Karami, H. (2024). A Rasch-based validation of the University of Tehran English Proficiency Test (UTEPT). *Language Testing in Asia, 14*, Article 18. https://doi.org/10.1186/s40468-024-00290-4

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, Article 100416. https://doi.org/10.1016/j.edurev.2021.100416

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education, 8,* Article 1217167. https://doi.org/10.3389/feduc.2023.1217167

Pupik Dean, C. G., Grossman, P., Enumah, L., Herrmann, Z., & Kavanagh, S. S. (2023). Core practices for project-based learning: Learning from experienced practitioners in the United States. *Teaching and Teacher Education, 133*, Article 104275. https://doi.org/10.1016/j.tate.2023.104275

Saky, S. A. T. M., Inayati, N. L., & Islam, M. N. (2025). Research patterns in formative assessment: A bibliometric review of primary and secondary school studies. *Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram, 13*(1), 107–122. https://doi.org/10.33394/j-ps.v13i1.14219

Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K–12 students' learning: Toward sustainable quality education. *Sustainability, 16*(17), Article 7826. https://doi.org/10.3390/su16177826

Suminar, I., Saputra, M. R. D., & Iqbal, N. H. M. (2025). Evaluating student-built earthquake alarm prototypes: Integrating engineering practices through real-world projects in junior high school. *Jurnal Pendidikan Fisika dan Teknologi, 11*(1a), 54–66. https://doi.org/10.29303/jpft.v11i1a.9306

Formatted: Font: Italic

Deleted: Washington, DC: U.S. Department of Education. Retrieved from

Deleted: Geelan, D., Hine, G., Ramirez-Campillo, R., . . . Xuan, Q.

Formatted: Font: Italic

Suyantiningsih, Badawi, Sumarno, Prihatmojo, A., Suprapto, I., & Munisah, E. (2023). Blended project-based learning (BPjBL) on students' achievement: A meta-analysis study. *International Journal of Instruction, 16*(3), 1113–1126. https://doi.org/10.29333/iji.2023.16359a

Tan, L. Y., McLean, S., Kim, Y. A., & Vitta, J. P. (2024). Rasch modelling vs. item facility: Implications for the validity of assessments of Asian EFL/ESL vocabulary knowledge and lexical sophistication modelling. *Language Testing in Asia, 14*, Article 55. https://doi.org/10.1186/s40468-024-00327-8

van der Steen, J., van Schilt-Mol, T., van der Vleuten, C., & Joosten-ten Brinke, D. (2023). Designing formative assessment that improves teaching and learning: What can be learned from the design stories of experienced teachers? *Journal of Formative Design in Learning, 7*(2), 182–194. https://doi.org/10.1007/s41686-023-00080-w

Woitt, S., Weidlich, J., Jivet, I., Orhan Göksün, D., Drachsler, H., & Kalz, M. (2025). Students' feedback literacy in higher education: An initial scale validation study. *Teaching in Higher Education, 30*(1), 257–276. https://doi.org/10.1080/13562517.2023.2263838

Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation, 74*, Article 101183. https://doi.org/10.1016/j.stueduc.2022.101183

Yao, Y., Amos, M., Snider, K., & Brown, T. (2024). The impact of formative assessment on K–12 learning: A meta-analysis. *Educational Research and Evaluation, 29*(7–8), 452–475. https://doi.org/10.1080/13803611.2024.2363831

Zhang, L., & Ma, Y. (2023). A study of the impact of project-based learning on student learning effects: A meta-analysis study. *Frontiers in Psychology, 14*, Article 1202728. https://doi.org/10.3389/fpsyg.2023.1202728

Zhao, L., Zhao, B., & Li, C. (2023). Alignment analysis of teaching–learning–assessment within the classroom: How teachers implement project-based learning under the curriculum standards. *Disciplinary and Interdisciplinary Science Education Research, 5*, Article 13. https://doi.org/10.1186/s43031-023-00078-1

Zulliger, S., Buholzer, A., & Ruelmann, M. (2022). Observed quality of formative peer and self-assessment in everyday mathematics teaching and its effects on student performance. *European Journal of Educational Research, 11*(2), 663–680. https://doi.org/10.12973/eu-jer.11.2.663