



PENGEMBANGAN TES BERPIKIR TINGKAT TINGGI BERBASIS QUIZZZ PADA ELEMEN DASAR-DASAR TEKNIK MESIN

C. Rudy Prihantoro¹, Hari Din Nugraha², Ahmad Fauzi³, Agung Gumelar⁴
Program Studi Pendidikan Teknik Mesin, Fakultas Teknik
Universitas Negeri Jakarta
e-mail: *crudy@unj.ac.id



ABSTRAK

Guru Dasar-Dasar Teknik Mesin di salah satu Sekolah Menengah Kejuruan (SMK) di Jakarta masih menyusun soal pada jenjang mengingat, memahami, dan menerapkan, padahal Kurikulum Merdeka menuntut asesmen yang melatih berpikir tingkat tinggi. Penelitian ini bertujuan menguraikan prosedur pengembangan, menguji validitas, dan menilai kepraktisan tes kognitif berpikir tingkat tinggi berbasis Quizizz untuk siswa Fase E. Penelitian pengembangan ini memakai desain sekuensial eksploratoris dengan tahap analisis, perancangan, pengembangan, dan implementasi. Dua puluh butir pilihan ganda pada jenjang menganalisis, mengevaluasi, dan mencipta divalidasi oleh ahli materi, ahli telaah soal, dan ahli bahasa, kemudian diujikan kepada 90 siswa Teknik Pemesinan dan dianalisis dengan model Rasch melalui program QUEST. Koefisien Aiken V validasi materi mencapai 0,847 (tinggi), validasi telaah soal 82,8%, validasi bahasa 100%, dan respons guru 90,9% (sangat praktis). Rerata Infit Mean Square sebesar 1,04 dan reliabilitas estimasi 0,89. Lima belas butir sesuai dengan model, tiga belas butir memiliki daya beda di atas 0,40, dan lima belas butir berada pada tingkat kesukaran sedang. Enam butir tidak berfungsi membedakan siswa, dan lima di antaranya merupakan butir mengevaluasi. Butir mengevaluasi dalam format pilihan ganda perlu ditelaah ulang, terutama pada pengecohnya, sebelum tes dipakai lebih luas.

Kata kunci: berpikir tingkat tinggi, instrumen penilaian, model Rasch, Quizizz, teknik mesin

ABSTRACT

Teachers of Basic Mechanical Engineering at a vocational high school in Jakarta still write test items mostly at the remembering, understanding, and applying levels, whereas the Merdeka Curriculum calls for assessment that trains higher-order thinking. This study aimed to describe the development procedure, examine the validity, and assess the practicality of a Quizizz-based higher-order thinking cognitive test for Phase E students. We used an exploratory sequential design within research and development, covering analysis, design, development, and implementation. Twenty multiple-choice items at the analyzing, evaluating, and creating levels were reviewed by content, item review, and language experts, administered to 90 machining students, and analyzed with the Rasch model in QUEST. Aiken's V for content validation reached 0.847 (high), the item review rating was 82.8%, the language rating 100%, and the teacher response 90.9% (very practical). The mean infit mean square was 1.04 and the reliability of estimate 0.89. Fifteen items fitted the model, thirteen items showed discrimination above 0.40, and fifteen items fell in the moderate difficulty range. Six items failed to discriminate, five of which were evaluating-level items. Evaluating-level multiple-choice items, particularly their distractors, need further review before the test is used more widely.

Keywords: *higher-order thinking, assessment instrument, Rasch model, Quizizz, mechanical engineering*

PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) menyiapkan lulusan untuk bekerja di industri, dan pekerjaan di bengkel pemesinan jarang berupa hafalan. Seorang operator mesin bubut harus membaca gambar kerja, menentukan urutan proses, menimbang risiko keselamatan, lalu memutuskan langkah perbaikan ketika ukuran benda kerja menyimpang. Keputusan semacam itu berada pada jenjang menganalisis, mengevaluasi, dan mencipta dalam taksonomi Bloom revisi (Anderson & Krathwohl, 2001), yang dalam kepustakaan pendidikan disebut kemampuan berpikir tingkat tinggi atau Higher Order Thinking Skills. Tanujaya et al. (2017) menemukan korelasi sebesar 0,814 antara skor berpikir tingkat tinggi dan indeks prestasi 41 mahasiswa pendidikan matematika. Hubungan sekuat itu menunjukkan bahwa kemampuan berpikir tingkat tinggi berjalan seiring dengan capaian akademik.

Kemampuan tersebut jarang dilatih dan jarang diukur di kelas. Retnawati et al. (2018) melaporkan bahwa pengetahuan guru tentang berpikir tingkat tinggi, kemampuan mereka melatihkannya kepada siswa, dan kemampuan mereka mengukurnya masih rendah. Dari sisi siswa, Hadi et al. (2018) mengkaji kesulitan siswa sekolah menengah dalam menyelesaikan soal berpikir tingkat tinggi. Abdullah et al. (2015) menelusuri jawaban 96 siswa dan menemukan bahwa kesalahan paling banyak terjadi pada tahap menuliskan jawaban akhir (27,58%) dan tahap keterampilan proses (27,33%). Siswa dalam penelitian itu kesulitan

menghubungkan informasi soal dengan strategi penyelesaian. Di pendidikan kejuruan, Sukatiman et al. (2020) menerapkan pembelajaran berbasis masalah dengan bantuan bertahap (scaffolding) untuk meningkatkan kemampuan berpikir tingkat tinggi siswa vokasi pada pembelajaran geomatika. Upaya pembelajaran semacam itu tetap membutuhkan alat ukur yang sahih agar hasilnya dapat dibuktikan.

Gejala yang sama terlihat di SMK tempat penelitian ini dilakukan. Observasi dan wawancara dengan guru Dasar-Dasar Teknik Mesin pada studi pendahuluan menunjukkan bahwa guru belum pernah menyusun instrumen berpikir tingkat tinggi dan belum pernah mengikuti pelatihan tentang hal itu. Soal diambil dari buku teks terbitan pemerintah dan sebagian besar berada pada jenjang mengingat (C1), memahami (C2), dan menerapkan (C3). Angket analisis kebutuhan yang diisi 20 siswa Fase E memperlihatkan bahwa 80% siswa kesulitan memahami materi Dasar-Dasar Teknik Mesin dan 80% menilai pembelajaran kurang menarik tanpa media pendukung. Seluruh siswa menyatakan membutuhkan latihan soal berpikir tingkat tinggi, dan 85% sangat setuju bahwa mereka memerlukan instrumen penilaian yang interaktif dan berbasis teknologi. Elemen Dasar-Dasar Teknik Mesin menjadi dasar bagi materi keselamatan kerja dan proses produksi yang dipelajari siswa pada fase berikutnya.

Platform kuis daring menawarkan cara menyajikan soal bertingkat tinggi dengan stimulus gambar dan umpan balik segera. Zhao (2019) melaporkan bahwa mahasiswa akuntansi menilai Quizizz

mudah digunakan dan menyenangkan, dan kelas yang lebih sering memakainya memberikan skor kepuasan lebih tinggi. Kajian Wang dan Tahir (2020) atas 93 penelitian tentang Kahoot!, platform kuis sejenis, menyimpulkan adanya pengaruh positif terhadap prestasi belajar, dinamika kelas, sikap, dan kecemasan siswa, walaupun sebagian penelitian tidak menemukan pengaruh berarti. Licorish et al. (2018) memperoleh gambaran serupa dari wawancara dengan 14 mahasiswa. Göksün dan Gürsoy (2019) memberi catatan penyeimbang: kegiatan yang digamifikasi dengan Quizizz menghasilkan kenaikan prestasi dan keterlibatan yang sedikit lebih rendah daripada kelompok kontrol. Tampilan yang menarik belum menjamin pengukuran yang baik, sehingga mutu butir soal tetap harus dibuktikan.

Di Indonesia, beberapa peneliti telah mengembangkan instrumen berpikir tingkat tinggi berbantuan Quizizz. Hamidah dan Wulandari (2021) memperoleh respons siswa rata-rata 89,5% pada instrumen yang mereka kembangkan, Nisa dan Pahlevi (2021) menyusun instrumen serupa untuk mata pelajaran kearsipan di SMK, dan Narassati et al. (2021) mengembangkan alat evaluasi Mekanika Teknik untuk SMK di DKI Jakarta dengan koefisien Aiken 0,823 untuk materi, 0,758 untuk instrumen, dan 0,833 untuk bahasa. Laporan Narassati et al. (2021) bertumpu pada penilaian ahli, sedangkan Hamidah dan Wulandari (2021) menekankan kelayakan dan respons siswa. Bukti kualitas butir dari model pengukuran modern, seperti kecocokan butir dengan model, tingkat kesukaran dalam satuan logit, dan daya beda, belum menjadi fokus kedua penelitian itu. Istiyono et al. (2014)

mengembangkan tes berpikir tingkat tinggi fisika untuk Sekolah Menengah Atas, kemudian Istiyono et al. (2019) membuktikan validitas empiris tes berpikir kritis fisika pada 252 siswa melalui statistik Infit Mean Square dalam kerangka teori respons butir. Boone (2016) menjelaskan bahwa analisis Rasch memungkinkan pengembang memeriksa fungsi pengukuran setiap butir dan memutuskan butir mana yang perlu diperbaiki. Penulis belum menemukan tes berpikir tingkat tinggi berbasis Quizizz untuk elemen Dasar-Dasar Teknik Mesin Kurikulum Merdeka yang diuji dengan cara tersebut.

Berdasarkan celah itu, penelitian ini bertujuan menguraikan prosedur pengembangan tes kognitif berpikir tingkat tinggi berbasis Quizizz pada elemen Dasar-Dasar Teknik Mesin untuk siswa Fase E, menguji validitasnya melalui penilaian ahli dan analisis model Rasch, serta menilai kepraktisannya menurut guru pengampu. Hasil penelitian diharapkan memberi guru SMK kumpulan butir yang telah teruji sekaligus menunjukkan jenjang kognitif yang paling sulit disusun dalam format pilihan ganda.

METODE

Penelitian ini merupakan penelitian dan pengembangan dengan desain sekuensial eksploratoris (Creswell & Creswell, 2018). Data kualitatif dari wawancara dan observasi dikumpulkan lebih dahulu untuk merumuskan kebutuhan produk, kemudian data kuantitatif dari validasi ahli dan uji coba dipakai untuk menilai produk. Pengembangan berlangsung dalam empat tahap, yaitu analisis, perancangan, pengembangan, dan implementasi. Penelitian dilaksanakan di salah satu

SMK di Jakarta yang memiliki Program Keahlian Teknik Pemесinan pada tahun ajaran 2024/2025.

Tahap analisis terdiri atas analisis kebutuhan dan analisis kurikulum. Wawancara dengan guru Teknik Pemесinan Fase E menggali sumber soal, bentuk soal, dan jenjang kognitif yang biasa digunakan, sedangkan angket tertutup diberikan kepada 20 siswa Fase E. Analisis kurikulum merujuk pada capaian pembelajaran Program Keahlian Dasar Fase E dalam Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan Nomor 033 Tahun 2022.

Tujuan pembelajaran yang semula belum memuat jenjang berpikir tingkat tinggi dirumuskan ulang oleh peneliti agar mencakup jenjang menganalisis (C4), mengevaluasi (C5), dan mencipta (C6).

Tahap perancangan menghasilkan kisi-kisi yang memuat elemen, capaian pembelajaran, tujuan pembelajaran, lingkup materi, jenjang kognitif, dan nomor butir. Kisi-kisi mencakup dua elemen, yaitu Keselamatan dan Kesehatan Kerja serta Lingkungan Hidup, dan Proses Produksi. Sebaran butir menurut elemen dan jenjang kognitif disajikan pada Tabel 1.

Tabel 1. Sebaran Butir Tes Menurut Elemen dan Jenjang Kognitif

Elemen	Lingkup materi	Menganalisis (C4)	Mengevaluasi (C5)	Mencipta (C6)	Jumlah butir
Keselamatan dan Kesehatan Kerja serta Lingkungan Hidup	Budaya kerja 5R (ringkas, rapi, resik, rawat, rajin), etika kerja, penanganan keadaan darurat, alat pelindung diri	1, 3	2, 4	-	4
Proses Produksi	Alat ukur, perkakas tangan dan bertenaga, mesin bubut, penggergajian, pemotongan, penggerindaan, pengelasan, pembentukan pelat, pengecoran logam	5, 6, 7, 11, 12, 16, 18, 19	8, 9, 10, 13, 14, 15, 20	17	16
Jumlah		10	9	1	20

Tahap pengembangan menghasilkan 20 butir pilihan ganda dengan lima pilihan jawaban. Setiap butir diawali stimulus berupa kasus atau gambar dari pekerjaan bengkel, lalu diunggah ke Quizizz dengan pengaturan waktu dan penskoran otomatis. Draf butir dinilai melalui empat jalur, yaitu validasi

materi oleh seorang dosen Pendidikan Teknik Mesin dan seorang guru Dasar-Dasar Teknik Mesin, validasi telaah soal oleh guru, validasi bahasa oleh dosen, dan angket respons guru untuk menilai kepraktisan. Seluruh penilaian ahli berlangsung pada 11 sampai 25 Juli 2025.

Tahap implementasi berupa uji coba kepada 90 siswa Fase E Program Keahlian Teknik Pemesinan. Kelas uji coba dipilih dengan teknik pengambilan sampel klaster secara acak (cluster random sampling), yaitu pemilihan kelompok secara acak dari kelompok-kelompok yang tersedia. Siswa mengerjakan tes melalui Quizizz, dan rekaman jawaban diekspor untuk dianalisis.

Validitas materi dihitung dengan koefisien Aiken V (Aiken, 1985). Dua validator memberi skor 1 sampai 5 pada sembilan aspek, antara lain kesesuaian instrumen dengan konsep Dasar-Dasar Teknik Mesin, kesesuaian dengan indikator dan tujuan pembelajaran, kejelasan batasan pertanyaan dan jawaban, kebenaran kalimat, ketersediaan petunjuk pengerjaan, dan kejelasan gambar. Koefisien dihitung dengan rumus $V = \sum s / [n(c - 1)]$, dengan s adalah selisih antara skor validator dan skor terendah, n adalah jumlah validator, dan c adalah skor tertinggi. Nilai $0 < V \leq 0,4$ dikategorikan rendah, $0,4 < V \leq 0,8$ sedang, dan $0,8 < V \leq 1$ tinggi.

Hasil telaah soal, validasi bahasa, dan angket respons guru diolah menjadi persentase, yaitu skor yang diperoleh dibagi skor maksimum lalu dikalikan 100%. Persentase 81,5% sampai 100% dikategorikan sangat valid, 62,5% sampai di bawah 81,5% valid, 43,5% sampai di bawah 62,5% kurang valid, dan 25% sampai di bawah 43,5% tidak valid. Untuk kepraktisan, persentase 86% sampai 100% berarti sangat praktis dan dapat digunakan tanpa revisi, 71% sampai 85% praktis, 56% sampai 70% cukup praktis, 41% sampai 55% kurang praktis, dan 25% sampai 40% tidak praktis.

Respons 90 siswa dianalisis dengan model Rasch menggunakan program QUEST. Butir dinyatakan sesuai dengan model apabila nilai Infit Mean Square berada pada rentang 0,5 sampai 1,5. Nilai di bawah 0,5 (overfit) menandakan pola jawaban yang terlalu mudah diduga, sedangkan nilai di atas 1,5 (misfit) menandakan pola jawaban yang menyimpang dari model. Tingkat kesukaran dikelompokkan menjadi mudah ($\text{logit} \leq -1,0$), sedang ($-1,0 < \text{logit} < +1,0$), dan sukar ($\text{logit} \geq +1,0$). Daya beda di atas 0,40 dikategorikan sangat baik, 0,30 sampai 0,39 baik, 0,20 sampai 0,29 cukup, 0,10 sampai 0,19 kurang, dan di bawah 0,10 tidak berfungsi. Konsistensi pengukuran dibaca dari reliabilitas estimasi (reliability of estimate) yang dilaporkan QUEST.

HASIL DAN PEMBAHASAN

Prosedur dan Produk Pengembangan

Analisis kebutuhan menghasilkan tiga temuan yang menjadi dasar perancangan. Guru belum pernah menyusun soal berpikir tingkat tinggi, soal yang digunakan berasal dari buku teks pemerintah dan sumber daring, dan jenjang kognitif yang dominan adalah C1 sampai C3. Guru juga menyadari bahwa tes yang selama ini diberikan belum memenuhi tuntutan Kurikulum Merdeka. Analisis kurikulum menunjukkan bahwa modul ajar Dasar-Dasar Teknik Mesin di sekolah belum memuat tujuan pembelajaran pada jenjang C4 sampai C6, sehingga peneliti merumuskan tujuan pembelajaran baru sebelum menyusun kisi-kisi.

Produk akhir berupa 20 butir pilihan ganda di Quizizz yang terdiri atas 10 butir menganalisis, 9 butir mengevaluasi, dan 1 butir mencipta (Tabel 1). Empat butir

mengukur elemen Keselamatan dan Kesehatan Kerja serta Lingkungan Hidup, dan 16 butir mengukur elemen Proses Produksi. Setiap butir dilengkapi gambar atau kasus bengkel sebagai stimulus. Stimulus ini ditambahkan setelah validator materi menyarankannya dalam lembar validasi.

Validitas Isi dan Kepraktisan

Hasil penilaian ahli dan guru dirangkum pada Tabel 2. Pada validasi materi, validator dosen memberi skor 43 dari 45 (95,5%) dan validator guru memberi skor 37 dari 45 (82,2%). Koefisien Aiken V per aspek berkisar 0,75 sampai 1, dengan rerata 0,847 yang

termasuk kategori tinggi. Validasi telaah soal memperoleh skor 29 atau 82,8%, dan validasi bahasa memperoleh skor 25 atau 100%. Keduanya termasuk kategori sangat valid dan tidak disertai permintaan revisi. Angket respons guru memperoleh skor 50 dari skor maksimum 55 atau 90,9%, yang termasuk kategori sangat praktis. Skor tertinggi diberikan guru pada keterkaitan soal dengan tujuan pembelajaran, tampilan gambar dan tulisan, kegunaan tes sebagai latihan mandiri, dan kemampuan tes melatih keterampilan berpikir. Kejelasan soal, daya dorong terhadap motivasi, dan daya tarik untuk mengerjakan soal memperoleh skor satu tingkat lebih rendah.

Tabel 2. Hasil Validasi Ahli dan Respons Guru

Jenis penilaian	Penilai	Skor	Hasil	Kategori
Validasi materi	Dosen Pendidikan Teknik Mesin dan guru Dasar-Dasar Teknik Mesin	43 dan 37 dari 45	Aiken V = 0,847	Tinggi
Validasi telaah soal	Guru Dasar-Dasar Teknik Mesin	29	82,8%	Sangat valid
Validasi bahasa	Dosen Pendidikan Teknik Mesin	25	100%	Sangat valid
Respons guru	Guru Dasar-Dasar Teknik Mesin	50 dari 55	90,9%	Sangat praktis

Kualitas Butir Berdasarkan Model Rasch

Rerata Infit Mean Square sebesar 1,04 dan rerata Outfit Mean Square sebesar 1,47; keduanya masih berada dalam rentang 0,5 sampai 1,5. Rerata nilai t Infit sebesar -0,12 dan rerata nilai t Outfit

sebesar -0,02. Tidak ada butir yang dijawab benar oleh seluruh siswa atau dijawab salah oleh seluruh siswa. Reliabilitas estimasi sebesar 0,89 termasuk kategori tinggi (0,71 sampai 0,90). Hasil analisis per butir disajikan pada Tabel 3.

Tabel 3. Hasil Analisis Butir dengan Model Rasch (n = 90)

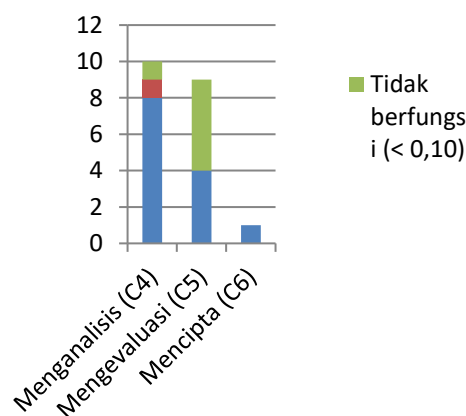
Butir	Jenjang	Infit Mean Square	Logit	Kesukaran	Daya beda	Keputusan
1	C4	0,30	-0,75	Sedang	0,94	Revisi (overfit)
2	C5	0,62	-0,28	Sedang	0,80	Dipertahankan
3	C4	1,95	-0,94	Sedang	0,08	Revisi
4	C5	2,14	-0,75	Sedang	-0,05	Revisi

5	C4	0,67	-0,15	Sedang	0,74	Dipertahankan
6	C4	0,98	1,14	Sukar	0,54	Dipertahankan
7	C4	0,81	0,60	Sedang	0,67	Dipertahankan
8	C5	0,68	0,60	Sedang	0,74	Dipertahankan
9	C5	2,06	-0,75	Sedang	0,03	Revisi
10	C5	0,87	0,32	Sedang	0,65	Dipertahankan
11	C4	0,73	0,10	Sedang	0,73	Dipertahankan
12	C4	0,81	0,32	Sedang	0,68	Dipertahankan
13	C5	0,69	0,10	Sedang	0,76	Dipertahankan
14	C5	1,75	-1,60	Mudah	-0,09	Revisi
15	C5	1,14	-3,43	Mudah	-0,05	Revisi
16	C4	0,58	-0,28	Sedang	0,79	Dipertahankan
17	C6	0,97	2,02	Sukar	0,43	Dipertahankan
18	C4	1,11	2,44	Sukar	0,32	Revisi ringan
19	C4	0,90	0,93	Sedang	0,60	Dipertahankan
20	C5	1,09	0,37	Sedang	0,05	Revisi

Keterangan: C4 = menganalisis; C5 = mengevaluasi; C6 = mencipta. Rentang Infit Mean Square yang diterima 0,5 sampai 1,5. Kesukaran: mudah ($\logit \leq -1,0$), sedang ($-1,0 < \logit < +1,0$), sukar ($\logit \geq +1,0$).

Lima belas butir memiliki Infit Mean Square di dalam rentang yang diterima. Butir 1 bernilai 0,30 (overfit), sedangkan butir 3, 4, 9, dan 14 bernilai 1,75 sampai 2,14 (misfit). Nilai logit berkisar dari -3,43 pada butir 15 sampai +2,44 pada butir 18. Dua butir (10%) tergolong mudah, 15 butir (75%) sedang, dan 3 butir (15%) sukar, yaitu butir 6, 17, dan 18.

Tiga belas butir memiliki daya beda di atas 0,40, dan butir 18 (0,32) berada pada kategori baik. Enam butir tidak berfungsi membedakan siswa karena daya bedanya di bawah 0,10, yaitu butir 3, 4, 9, 14, 15, dan 20; empat di antaranya juga misfit. Bila dikelompokkan menurut jenjang kognitif (Gambar 1), butir yang tidak berfungsi terkumpul pada jenjang mengevaluasi: lima dari sembilan butir C5, dibandingkan satu dari sepuluh butir C4. Satu-satunya butir C6, yaitu butir 17, memiliki daya beda 0,43 dan tergolong sukar.



Gambar 1. Sebaran Daya Beda Butir Menurut Jenjang Kognitif

Pembahasan

Prosedur empat tahap menghasilkan tes yang dinilai layak oleh ahli, dan sebagian besar butirnya berfungsi ketika dijawab siswa. Koefisien Aiken V validasi materi sebesar 0,847 sedikit lebih tinggi daripada koefisien validasi materi 0,823 yang dilaporkan Narassati et al. (2021) untuk alat evaluasi Mekanika Teknik berbasis Quizizz di SMK DKI Jakarta. Kedua angka berada pada kategori tinggi, tetapi perbandingan ini perlu dibaca hati-hati karena validator,

jumlah aspek, dan skala penilaiannya berbeda. Penilaian ahli pada kedua penelitian juga baru menjawab apakah butir tampak sesuai dengan indikator. Penilaian itu belum memperlihatkan bagaimana butir bekerja ketika dihadapkan pada siswa dengan kemampuan yang beragam.

Analisis Rasch memberi informasi yang tidak tersedia dari penilaian ahli. Reliabilitas estimasi 0,89 lebih tinggi daripada reliabilitas 0,75 yang dicapai Istiyono et al. (2019) pada tes berpikir kritis fisika berbasis komputer, meskipun konstruk, jumlah butir, dan model analisis kedua penelitian berbeda. Rerata Outfit Mean Square sebesar 1,47 berada dekat batas atas rentang yang diterima. Statistik outfit peka terhadap jawaban yang tidak terduga, misalnya siswa berkemampuan tinggi yang salah pada butir mudah atau siswa berkemampuan rendah yang menebak benar pada butir sukar. Nilai yang mendekati batas ini sejalan dengan adanya empat butir misfit dan enam butir yang tidak berfungsi membedakan siswa. Boone (2016) menempatkan penemuan butir semacam ini sebagai salah satu kegunaan utama analisis Rasch dalam pengembangan instrumen.

Temuan yang paling menonjol adalah terkumpulnya butir yang tidak berfungsi pada jenjang mengevaluasi. Lebih dari separuh butir C5 gagal membedakan siswa berkemampuan tinggi dan rendah, sementara hampir semua butir C4 berfungsi. Butir 14 dan 15 memperlihatkan pola yang paling jelas: keduanya tergolong mudah, tetapi daya bedanya negatif. Pola ini terjadi ketika siswa berkemampuan rendah memilih kunci jawaban sama seringnya atau lebih sering daripada siswa berkemampuan tinggi. Butir 14 meminta siswa

menentukan langkah penyelesaian pada proses pemotongan, dan butir 15 meminta siswa menangani masalah pada proses penggerindaan secara aman dan efisien. Kedua tugas itu menuntut penilaian berdasarkan kriteria, dan dalam format pilihan ganda kriteria tersebut harus tertanam pada pilihan jawaban. Bila lebih dari satu pilihan dapat dibenarkan dari sudut pandang praktik bengkel, siswa yang lebih menguasai materi justru dapat terpecah pilihannya. Penjelasan ini masih berupa dugaan penulis dan perlu diuji dengan menelaah ulang sebaran pilihan jawaban setiap butir C5 bersama guru dan ahli materi.

Temuan tersebut sejalan dengan laporan Retnawati et al. (2018) bahwa kemampuan guru mengukur berpikir tingkat tinggi masih rendah. Dalam penelitian ini, butir disusun oleh peneliti dan ditelaah oleh guru serta dosen, tetapi kelemahan butir C5 baru terlihat setelah analisis empiris. Penilaian ahli yang tinggi secara keseluruhan tidak menangkap kelemahan tersebut. Guru yang menyusun soal mengevaluasi tanpa uji coba berpeluang menghadapi masalah yang sama.

Butir 1 memperlihatkan persoalan yang berbeda. Daya bedanya paling tinggi (0,94), tetapi Infit Mean Square hanya 0,30 dan 90% siswa memilih kunci jawaban. Pola jawaban yang terlalu mudah diduga membuat butir ini memberi sedikit informasi tambahan tentang kemampuan siswa, sehingga butir tersebut masih layak dipakai setelah pengecohnya diperkuat. Di ujung yang lain, butir 17 dan 18 merupakan butir tersukar. Butir 17, yang meminta siswa menyusun alur kerja proses pengelasan, tetap berfungsi baik walaupun sukar. Karena jenjang mencipta hanya diwakili

satu butir, temuan ini belum dapat digeneralisasi untuk jenjang C6.

Dari sisi kepraktisan, guru memberi skor tertinggi pada tampilan, gambar, dan kegunaan tes sebagai latihan mandiri. Temuan ini sejalan dengan Zhao (2019) serta respons siswa 89,5% yang dilaporkan Hamidah dan Wulandari (2021). Penskoran otomatis di Quizizz juga mempercepat guru memperoleh hasil tes. Meski demikian, Göksün dan Gürsoy (2019) mengingatkan bahwa daya tarik platform tidak selalu diikuti kenaikan prestasi. Penelitian ini mengukur kepraktisan dari satu guru dan belum mengukur perubahan kemampuan berpikir tingkat tinggi siswa, sehingga klaim bahwa tes ini meningkatkan kemampuan tersebut belum dapat diajukan.

Guru Dasar-Dasar Teknik Mesin dapat langsung memakai dua belas butir yang sesuai dengan model dan berdaya beda sangat baik, lalu memperbaiki delapan butir lainnya, terutama butir mengevaluasi, dengan menyusun pengecoh dari kesalahan yang nyata dilakukan siswa di bengkel. Keterbatasan penelitian ini mencakup uji coba di satu sekolah dengan 90 siswa, validasi materi oleh dua orang, telaah soal dan validasi bahasa yang masing-masing oleh satu orang, serta penilaian kepraktisan yang hanya berasal dari satu guru. Penelitian lanjutan perlu menguji ulang butir yang telah direvisi pada beberapa SMK, menambah butir C6, dan memakai desain prates dan pascates untuk mengetahui pengaruh latihan dengan tes ini terhadap kemampuan berpikir tingkat tinggi siswa.

KESIMPULAN

Pengembangan tes kognitif berpikir tingkat tinggi berbasis Quizizz untuk

elemen Dasar-Dasar Teknik Mesin dapat dilakukan melalui analisis kebutuhan dan kurikulum, perumusan ulang tujuan pembelajaran ke jenjang menganalisis, mengevaluasi, dan mencipta, penyusunan kisi-kisi dan butir bersituasi bengkel, validasi ahli, serta uji coba yang dianalisis dengan model Rasch. Tes yang dihasilkan valid menurut ahli materi, ahli telaah soal, dan ahli bahasa, memiliki reliabilitas tinggi, dan sebagian besar butirnya sesuai dengan model serta mampu membedakan siswa berkemampuan tinggi dan rendah. Guru menilai tes ini sangat praktis untuk dipakai di kelas. Kelemahan tes terkumpul pada butir mengevaluasi, yang lebih sering gagal membedakan kemampuan siswa daripada butir menganalisis. Butir mengevaluasi perlu diperbaiki, terutama pada pengecohnya, dan diuji ulang pada sampel yang lebih luas sebelum tes dipakai sebagai alat ukur baku di SMK.

DAFTAR PUSTAKA

- Abdullah, A. H., Abidin, N. L. Z., & Ali, M. (2015). Analysis of students' errors in solving higher order thinking skills (HOTS) problems for the topic of fraction. *Asian Social Science*, 11(21), 133–142. <https://doi.org/10.5539/ass.v11n21.p133>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, 15(4), rm4. <https://doi.org/10.1187/cbe.16-04-0148>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Göksün, D. O., & Gürsoy, G. (2019). Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz. *Computers & Education*, 135, 15–29. <https://doi.org/10.1016/j.compedu.2019.02.015>
- Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. *Problems of Education in the 21st Century*, 76(4), 520–532. <https://doi.org/10.33225/pec/18.76.520>
- Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan instrumen penilaian berbasis HOTS menggunakan aplikasi “Quizizz”. *Efisiensi: Kajian Ilmu Administrasi*, 18(1), 105–124. <https://doi.org/10.21831/efisiensi.v18i1.36997>
- Istiyono, E., Dwandaru, W. S. B., Lede, Y. A., Rahayu, F., & Nadapdap, A. (2019). Developing IRT-based physics critical thinking skill test: A CAT to answer 21st century challenge. *International Journal of Instruction*, 12(4), 267–280. <https://doi.org/10.29333/iji.2019.12417a>
- Istiyono, E., Mardapi, D., & Suparno. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (PysTHOTS) peserta didik SMA. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 1–12. <https://doi.org/10.21831/pep.v18i1.2120>
- Licorish, S. A., Owen, H. E., Daniel, B., & George, J. L. (2018). Students’ perception of Kahoot!’s influence on teaching and learning. *Research and Practice in Technology Enhanced Learning*, 13, Article 9. <https://doi.org/10.1186/s41039-018-0078-8>
- Narassati, N. A., Saleh, R., & Arthur, R. (2021). Pengembangan alat evaluasi berbasis HOTS menggunakan aplikasi Quizizz pada mata pelajaran mekanika teknik dalam pembelajaran jarak jauh. *Jurnal Pendidikan Teknik Sipil*, 3(2), 169–180. <https://doi.org/10.21831/jpts.v3i2.43919>
- Nisa, S., & Pahlevi, T. (2021). Pengembangan instrumen penilaian HOTS berbantuan Quizizz pada mata pelajaran kearsipan SMK. *Edukatif: Jurnal Ilmu Pendidikan*, 3(5), 2146–2159. <https://doi.org/10.31004/edukatif.v3i5.756>
- Retnawati, H., Djidu, H., Kartianom, Apino, E., & Anazifa, R. D. (2018). Teachers’ knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century*, 76(2), 215–230. <https://doi.org/10.33225/pec/18.76.215>
- Sukatiman, Akhyar, M., Siswandari, & Roemintoyo. (2020). Enhancing

- higher-order thinking skills in vocational education through scaffolding-problem based learning. *Open Engineering*, 10(1), 612–619. <https://doi.org/10.1515/eng-2020-0070>
- Tanujaya, B., Mumu, J., & Margono, G. (2017). The relationship between higher order thinking skills and academic performance of student in mathematics instruction. *International Education Studies*, 10(11), 78–85. <https://doi.org/10.5539/ies.v10n11p78>
- Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning – A literature review. *Computers & Education*, 149, 103818. <https://doi.org/10.1016/j.compedu.2020.103818>
- Zhao, F. (2019). Using Quizizz to integrate fun multiplayer activity in the accounting classroom. *International Journal of Higher Education*, 8(1), 37–43. <https://doi.org/10.5430/ijhe.v8n1p37>