

ANALISIS KUALITAS BUTIR SOAL TIPE HOTS PADA KOMPETENSI SISTEM REM SISWA DI SEKOLAH MENENGAH KEJURUAN

Warju¹⁾, Sudirman Rizki Ariyanto²⁾, Soeryanto³⁾, Rio Adi Trisna⁴⁾

^{1,2,3,4} Jurusan Teknik Mesin, Fakultas Teknik, Universitas Negeri Surabaya
Email: warju@unesa.ac.id, sudirmanariyanto@mhs.unesa.ac.id, soeryanto@unesa.ac.id,
riotrisna@mhs.unesa.ac.id

ABSTRAK

Tujuan dari penelitian ini adalah untuk mengetahui kualitas dari soal tipe HOTS pada kompetensi sistem rem melalui pelaksanaan analisis butir soal. Penelitian ini merupakan penelitian evaluatif. Sampel penelitian ini adalah siswa kelas XI TKRO 1 di SMKN 1 Labang Bangkalan sejumlah 32 siswa. Instrumen penelitian ini meliputi soal pilihan ganda tipe HOTS, kunci jawaban, dan lembar jawaban hasil pekerjaan siswa, yang kemudian dianalisis validitas, reliabilitas tingkat kesukaran, daya pembeda, dan efektivitas pengecoh menggunakan bantuan software Anates V4. Hasil penelitian ini menunjukkan bahwa dari total 50 soal pilihan ganda terdapat 80% soal yang masuk dalam kategori valid, sedangkan 20% soal masuk pada kategori tidak valid. Kemudian untuk reliabilitas mendapatkan nilai koefisien sebesar 0,917. Sementara itu soal yang masuk kategori mudah, sedang, dan sukar pada indeks tingkat kesukaran masing-masing sebesar 38%, 46%, dan 16%. Kemudian soal yang masuk kategori baik sekali, baik, cukup, buruk, dan tidak baik pada indeks daya beda masing-masing sebesar 10%, 42%, 28%, 6%, dan 14%. Selanjutnya pada indeks pengecoh soal yang masuk kriteria sangat baik, baik, kurang baik, buruk, dan sangat buruk masing-masing sebesar 28%, 29%, 19%, 17%, dan 7%. Dengan demikian dapat disimpulkan bahwa soal tipe HOTS pada kompetensi sistem rem layak digunakan sebagai alat penilai kompetensi siswa.

Kata kunci: soal pilihan ganda tipe HOTS, validitas, reliabilitas, tingkat kesukaran, daya pembeda, tingkat kesukaran, SMKN 1 Labang Bangkalan, kompetensi sistem rem

ABSTRACT

The purpose of this study is to study the quality of HOTS questions on brake system competency through the item analysis. This research is an evaluative study. The sample of this research is 32 students of TKRO 1 in SMK Negeri 1 Labang Bangkalan with a total of 32 students. HOTS, answer keys, and answer sheets of student work, which are then checked for validity, reliability level of compatibility, distinguishing power, and deception effectiveness using the help of Anates V4 software. The results of this study indicate that from a total of 50 multiple choice questions there are 80% of the items included in the valid category, while 20% of the items belong to the invalid category. Then for reliability, get a coefficient of 0.917. Meanwhile, the items were categorized as easy, medium, and difficult in the suitability index of 38%, 46%, and 16%, respectively. Then the items that are categorized as very good, good, enough, bad, and not good at different power index each of 10%, 42%, 28%, 6%, and 14%. Then in the index check the questions that came in were very good, good, not good, bad, and very bad each of 28%, 29%, 19%, 17%, and 7%. Thus, it can dispute the matter of the type of HOTS on the brake system competency fit to be used as a student competency assessment tool.

Keywords : HOTS multiple-choice questions, validity, reliability, level of difficulty, differentiating power, level of difficulty, SMKN 1 Labang Bangkalan, brake system competency

1. PENDAHULUAN

Siswa sekolah menengah kejuruan umumnya memperoleh pengetahuan dan keterampilan melalui serangkaian proses pembelajaran di sekolah. Khusus di Sekolah Menengah Kejuruan proses pembelajaran dibagi menjadi dua bagian yakni pembelajaran secara teori yang dilakukan di kelas dan pembelajaran secara praktik yang dilakukan di bengkel sekolah. Setelah serangkaian pembelajaran

telah selesai dilakukan selanjutnya guru melakukan penilaian untuk mengukur sejauh mana siswa mampu mengembangkan kemampuannya sesuai dengan apa yang telah dipelajari. Oleh karena itu dalam melakukan penilaian sangat dibutuhkan alat atau instrumen yang valid dan mampu mengukur kemampuan siswa secara objektif [1], [2]. Diantara berbagai macam alat penilaian yang bisa digunakan, soal pilihan ganda menjadi alat ukur yang paling banyak digunakan untuk penilaian siswa [3], [4]. Hal tersebut disebabkan oleh luasnya materi yang dapat dicakup dalam soal pilihan ganda dengan waktu penilaian yang relatif singkat. Dengan demikian bias yang diperoleh dari hasil penilaian dapat diminimalisirkan.

Apabila disusun dengan tepat, soal pilihan ganda dapat digunakan untuk mengukur kemampuan *higher order of thinking skill* (HOTS) yang meliputi kemampuan menerapkan, menganalisis, mengevaluasi, dan mencipta [5], [6]. Umumnya, soal pilihan ganda terdiri dari empat pilihan jawaban. Dari empat jawaban tersebut terdapat satu jawaban benar dan tiga diantaranya salah. Tiga jawaban salah tersebut sering disebut sebagai “pengecoh” karena dapat mengalihkan siswa dari pemilihan jawaban yang benar. Soal pilihan ganda yang valid umumnya terdiri dari butir-butir soal yang berkualitas. Hal tersebut ditentukan melalui hasil uji validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas pengecoh [7]. Dalam implementasinya validitas butir soal dilakukan apabila peneliti telah menyelesaikan tahap validasi isi dan melakukan uji coba lapangan [8]. Kemudian dilakukan uji reliabilitas yang bertujuan untuk mengukur keajegan dari butir soal pilihan ganda. Keajegan dalam hal ini mengarah kepada kestabilan, kehandalan, dan kemampuan *predictability* suatu soal pilihan ganda yang telah disusun [9]. Salah satu teknik yang dapat digunakan dalam mengukur reliabilitas soal pilihan ganda adalah Teknik *Cronbach's Alpha*. Teknik ini memiliki tiga kategori, yakni: (1) jika $\alpha < 0,7$ maka butir soal masuk kategori kurang meyakinkan; (2) jika $\alpha \geq 0,7$ maka butir soal masuk kategori baik; dan (3) jika $\alpha \geq 0,8$ maka butir soal masuk kategori istimewa [10], [11].

Selanjutnya yakni uji tingkat kesukaran. Dalam hal ini uji tingkat kesukaran berhubungan dengan keseimbangan proporsi soal pilihan ganda baik yang masuk kategori mudah, sedang, maupun sukar. Menurut Sudjana (2017) proporsi tingkat kesukaran dapat menggunakan perbandingan 3-4-3, yang berarti terdapat 30% butir soal dengan kategori mudah, 40% butir soal dengan kategori sedang, dan 30% butir soal dengan kategori sukar. Selain itu, proporsi tingkat kesukaran juga dapat disusun dengan perbandingan 3-5-2, yang berarti terdapat 30% butir soal dengan kategori mudah, 50% butir soal dengan kategori sedang, dan 20% butir soal dengan kategori sukar [12]. Sementara itu, Rao, Kishan Prasad, Sajitha, Permi, & Shetty (2016) menyebutkan bahwa proporsi tingkat kesukaran juga dapat mengacu pada perbandingan 30% butir soal masuk kategori sukar dan 70% butir soal masuk kategori dapat diterima [13]. Adapun klasifikasi indeks tingkat kesukaran yang digunakan sebagai acuan meliputi: (1) jika butir soal berada pada interval 0,00 – 0,29 maka butir soal masuk pada kategori sukar; (2) jika butir soal berada pada interval 0,30-0,70 maka butir soal masuk pada kategori sedang; dan (3) jika butir soal berada pada interval 0,70-1,00 maka butir soal masuk pada kategori mudah.

Setelah uji tingkat kesukaran, maka selanjutnya dilakukan analisis daya pembeda. Tujuan dilakukan analisis ini adalah mengkaji kemampuan soal pilihan ganda dalam membedakan siswa dengan kategori berprestasi tinggi dan siswa dengan kategori berprestasi rendah. Daya pembeda dapat diketahui melalui indeks daya pembeda yang meliputi: (1) $0,7 < D \leq 1$ dengan kategori baik sekali; (2) $0,4 < D \leq 0,7$ dengan kategori baik; (3) $0,2 < D \leq 0,4$ dengan kategori cukup; (4) $0 < D \leq 0,2$ dengan kategori buruk; dan (5) $D \leq 0$ dengan kategori soal tidak baik [14], [15]. Tahap akhir dari analisis butir soal adalah analisis efektivitas pengecoh. Pengecoh merupakan pilihan jawaban yang disediakan oleh seorang peneliti dengan tujuan untuk menyesatkan (mengecoh) siswa yang kurang begitu paham terhadap materi yang diujikan. Pengecoh dapat dikatakan berfungsi dengan baik apabila terdapat minimal 5% siswa yang memilih pengecoh tersebut. Pengecoh yang terbukti tidak efektif untuk digunakan, maka disarankan untuk mengganti pengecoh tersebut dengan pengecoh yang lebih baik [16].

Pada dasarnya melakukan analisis butir soal sangat penting untuk dilakukan oleh guru. Hal tersebut bertujuan agar guru dapat mengetahui kehandalan dan validitas dari soal yang telah mereka susun [17]. Selain itu, melalui analisis butir soal akan mendapatkan umpan balik yang nantinya dapat dijadikan acuan dalam pengambilan keputusan. Soal yang masuk kategori layak dapat langsung digunakan, soal dengan kategori kurang layak dapat direvisi, sedangkan soal dengan kategori tidak layak dapat digugurkan sehingga harus dihapus atau diganti dengan soal lainnya [18], [19]. Namun sayangnya, pengetahuan dan keterampilan guru dalam melakukan analisis butir soal masih tergolong rendah [20]. Sebagian besar guru cenderung membuat soal sesuai dengan kompetensi inti dan kompetensi dasar yang telah ditentukan tanpa mengukur terlebih dahulu apakah siswa telah

memahami materi-materi yang akan diujikan. Dengan demikian sudah dapat dipastikan bahwa kompetensi siswa tentunya tidak dapat diukur secara tepat [21]

Tolak ukur dari keberhasilan pelaksanaan analisis butir soal dapat diketahui berdasarkan hasil penelitian sebelumnya. Rao, Kishan Prasad, Sajitha, Permi, & Shetty (2016) menemukan bahwa sebagian besar butir soal yang dikembangkan telah memenuhi kriteria kesukaran yang dapat diterima dan memiliki daya pembeda yang baik. Butir soal yang baik menunjukkan indeks daya pembeda yang buruk dan butir soal yang mudah menunjukkan daya pembeda yang tinggi. Hasil penelitian ini akan memulai perubahan dalam cara penyusunan soal-soal tes (*multiple choice question*) MCQ yang dipilih dalam setiap ujian dan harus ada strategi penilaian yang tepat sebagai bagian dari pengembangan kurikulum [13]. Kaur, Singla, & Mahajan (2016) melalui hasil penelitiannya menyimpulkan bahwa analisis butir soal merupakan alat yang berharga karena dapat membantu guru untuk mempertahankan soal pilihan ganda yang valid dan membuang butir-butir soal yang tidak layak digunakan. Hal ini juga membantu dalam meningkatkan keterampilan guru dalam menyusun dan mengidentifikasi butir-butir soal yang penting untuk diujikan kepada siswa [22]. Selanjutnya Brown & Abdulnabi (2017) melalui hasil penelitiannya menunjukkan bahwa penggunaan analisis item IRT memiliki dampak menguntungkan yang potensial pada keseluruhan nilai dan jumlah siswa yang lulus. Ini juga menunjukkan bahwa umpan balik sangat bermanfaat bagi guru sebagai acuan dalam penyusunan soal yang sesuai dengan karakteristik [4].

Quaigrain & Arhin (2017) melalui hasil penelitiannya menyimpulkan bahwa menyusun soal pilihan ganda untuk ujian akhir semester membutuhkan waktu dan harus diimbangi dengan pemilihan materi yang tepat. Kontrol kualitas penting untuk pengembangan soal. Oleh karena itu, soal dengan indeks diskriminasi negatif harus dihapus atau diganti. Guru kelas harus menulis ulang semua soal dengan nol indeks diskriminasi. Selain itu, guru harus mengganti atau menulis ulang semua soal dengan indeks diskriminasi positif rendah. Ketika semua ini dilakukan, reliabilitas tes akan meningkat. [23]. Danuwijaya (2018) dalam hasil penelitiannya menyebutkan bahwa sebagian besar butir soal rata-rata masuk kategori sukar. Dalam hal daya beda, lebih dari setengah dari total butir soal dikategorikan cukup sehingga memerlukan modifikasi lebih lanjut [24]. Sementara itu Kusumawati & Hadi (2018) menemukan bahwa dari 353 siswa yang menjadi sampel 40% dari 35 item yang diuji masuk kategori sangat sulit, 60% berada di level sedang, dan tidak ada item yang mudah. Persentase indeks diskriminasi item untuk sangat rendah, rendah, sedang, tinggi, dan sangat tinggi masing-masing adalah 8,57%, 51,43%, 31,43%, 5,71%, dan 2,86%. Selain itu distraktor pada semua butir soal masuk pada kategori berfungsi dengan baik [25].

Merujuk dari uraian pendahuluan dan beberapa hasil penelitian dahulu maka diketahui bahwa pelaksanaan analisis butir soal sangat perlu untuk dilakukan oleh guru. Oleh karena itu dalam penelitian ini dilakukan sebuah analisis butir soal tipe HOTS pada kompetensi sistem rem siswa SMK. Dengan demikian tujuan dari penelitian ini adalah untuk menganalisis hasil validasi, validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal tipe HOTS pada kompetensi sistem rem.

2. METODE

Penelitian ini merupakan penelitian evaluatif melalui serangkaian kegiatan pengumpulan data hasil penilaian, dimana kemudian hasil tersebut dibandingkan dengan kriteria yang telah ditentukan. Soal tipe HOTS yang digunakan terdiri dari 50 soal berupa pilihan ganda pada kompetensi sistem rem. Sampel penelitian ini adalah siswa kelas XI TKRO 1 di SMKN 1 Labang Bangkalan yang sedang mempelajari kompetensi sistem rem pada tahun ajaran 2019/2020 sejumlah 32 siswa. Semua standar dan etika pengumpulan data dipatuhi secara ketat. Teknik pengumpulan data yang digunakan berupa dokumentasi. Instrumen penelitian ini meliputi soal pilihan ganda tipe HOTS, kunci jawaban, dan lembar jawaban hasil pekerjaan siswa. Pelaksanaan tes dilakukan selama 60 menit. Hasil penelitian dianalisis validitas, reliabilitas tingkat kesukaran, daya pembeda, dan efektivitas pengecoh menggunakan bantuan *software* Anates V4 dengan taraf signifikansi sebesar 5%.

3. HASIL DAN PEMBAHASAN

A. Validasi Ahli (*Expert Judgement*)

Dalam penelitian analisis butir soal pilihan ganda tipe HOTS diawali dengan uji validasi dari para ahli yang terdiri dari dua orang dosen jurusan teknik mesin Universitas Negeri Surabaya dan satu orang guru dari jurusan teknik kendaraan ringan otomotif (TKRO) SMKN 1 Labang Bangkalan.

Pelaksanaan validasi dibagi dalam tiga aspek penilaian yang meliputi: (1) materi; (2) konstruksi; dan (3) bahasa atau budaya. Adapun hasil validasi dapat dilihat pada Tabel 1.

Tabel 1. Hasil Validasi soal pilihan ganda tipe HOTS

No	Aspek	Rata-rata	Kategori
1.	Materi	4,54	Sangat Valid
2.	Konstruksi	4,62	Sangat Valid
3.	Bahasa atau Budaya	4,65	Sangat Valid
Rata-rata Total		4,60	Sangat Valid

Berdasarkan data pada Tabel 1 didapatkan hasil validasi rata-rata dari ketiga validator sebesar 4,60 dengan kategori sangat valid. Hasil validasi tersebut menunjukkan bahwa soal pilihan ganda tipe HOTS memiliki kesesuaian baik dari aspek materi, konstruksi, maupun bahasa atau budaya. Oleh karena itu, dapat dinyatakan bahwa soal tersebut layak apabila digunakan untuk mengukur atau mengetahui hasil belajar siswa khususnya pada kompetensi sistem rem. Hasil ini sejalan dengan penelitian Sa'idah, Yulistianti, & Megawati (2018) yang menunjukkan bahwa sebelum melakukan analisis butir soal melalui uji coba lapangan, butir-butir soal terlebih dahulu dianalisis berdasarkan hasil validasi para ahli (*expert judgement*). Hal tersebut bertujuan untuk melihat kualitas produk awal dari butir soal pilihan ganda tipe HOTS yang kemudian disempurnakan berdasarkan masukan dan saran berdasarkan hasil *review* [26]

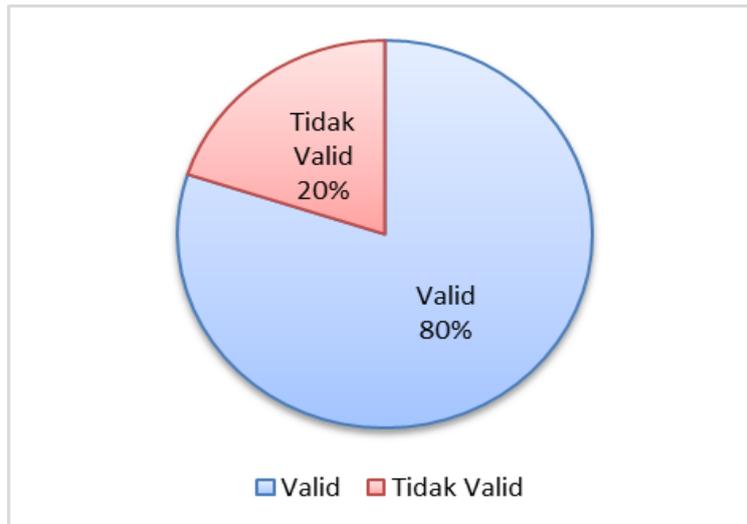
B. Validitas

Validitas menjelaskan tentang seberapa baik data yang dikumpulkan dapat mencakup area aktual yang sedang diteliti [27]. Pada penelitian ini hasil analisis validitas butir soal diuji dengan taraf signifikansi 5%, sehingga karena jumlah (N) sebanyak 32 siswa maka nilai r_{tabel} product moment sebesar 0,349. Berdasarkan hasil analisis, butir soal dapat dikategorikan valid apabila nilai $Y_{pbi} > 0,349$ dan soal dikategorikan tidak valid apabila nilai $Y_{pbi} < 0,349$ [28]. Adapun distribusi soal yang dibagi berdasarkan indeks validitas dapat dilihat pada Tabel 2 berikut.

Tabel 2. Distribusi soal berdasarkan indeks validitas

No.	Indeks Validitas	No. Butir Soal	Jumlah	Persentase
1.	Soal Valid ($Y_{pbi} \geq 0,349$)	1, 2, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 31, 32, 33, 34, 36, 37, 38, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50	40	80%
2.	Soal Tidak Valid ($Y_{pbi} < 0,349$)	3, 4, 5, 11, 15, 26, 30, 35, 39, 44	10	20%
Jumlah Total			50	100%

Dari data pada Tabel 1 di atas, apabila ditampilkan dalam bentuk diagram lingkaran atau diagram pie, maka nampak seperti pada Gambar 1 berikut ini.



Gambar 1. Diagram *pie* distribusi soal berdasarkan indeks validitas

Berdasarkan gambar Gambar 1, diketahui bahwa dari total 50 soal pilihan ganda terdapat 40 soal yang masuk dalam kategori “Valid” dengan persentase sebesar 80%, sedangkan 10 soal masuk pada kategori “Tidak Valid” dengan persentase sebesar 20%. Sebagaimana yang dinyatakan oleh Souza et al. (2017) bahwa validitas umumnya mengacu pada fakta bahwa alat ukur memiliki ketepatan atau keakurasian dalam mengukur apa yang hendak diukur [29]. Selain itu Setiawaty, Sulistyorini, Margono, & Rahmawati (2018) juga menjelaskan bahwa suatu instrumen dapat digunakan jika instrumen tersebut mampu menghasilkan hasil yang sama untuk mengevaluasi suatu pengukuran [30]. Merujuk dari beberapa pendapat dan hasil uji validitas, maka dapat dinyatakan bahwa sejumlah 40 butir soal pilihan ganda tipe HOTS untuk kompetensi sistem rem telah memenuhi beberapa unsur dan layak digunakan karena masuk dalam kategori “Valid”.

C. Reliabilitas

Reliabilitas merupakan sejauh mana butir soal mampu menghasilkan hasil yang sama pada uji coba secara berulang. Singkatnya, reliabilitas berkaitan stabilitas atau konsistensi skor dari waktu ke waktu atau lintas penilai. Perlu diingat bahwa keandalan berkaitan dengan skor, bukan pada orang [31]. Hasil pengolahan data kehandalan butir soal hasil belajar ranah kognitif dilakukan menggunakan teknik *Cronbach's Alpha*. Sementara itu, hasil uji reliabilitas butir soal tipe HOTS pada kompetensi sistem rem dapat dilihat pada Tabel 3 berikut.

Tabel 3. Hasil uji reliabilitas butir soal

Reliability Statistics	
Cronbach's Alpha	N of Items
0,917	50

Dari hasil uji pada Tabel 3 di atas, diketahui bahwa butir soal tipe HOTS pada kompetensi sistem rem mendapatkan nilai koefisien reliabilitas sebesar 0,917. Nilai tersebut menunjukkan bahwa nilai $\alpha \geq 0,8$ yang berarti bahwa butir soal masuk kategori istimewa [10]. Dengan demikian dinyatakan bahwa soal pilihan ganda tipe HOTS pada kompetensi sistem rem layak digunakan dan dapat dipercaya. Hasil penelitian ini sejalan dengan Al-Osail et al. (2015) yang menyimpulkan bahwa dari tiga kelompok yang mengikuti ujian OSCE dan tertulis didapatkan nilai *Cronbach's Alpha* untuk setiap kelompok adalah 0,7, 0,8, dan 0,9. Hasil ini menunjukkan stabilitas dan konsistensi internal yang baik dikarenakan perbedaan nilai *Cronbach's Alpha* diantara ketiga kelompok tersebut relatif kecil [32].

D. Tingkat Kesukaran

Soal tipe HOTS pada kompetensi sistem rem dapat dikategorikan baik apabila soal tersebut tidak terlalu mudah atau tidak terlalu sukar [33]. Untuk mengetahui tingkat kesukaran pada masing-masing butir soal, seorang guru perlu melakukan suatu pengujian. Tingkatan kesukaran pada setiap butir soal dibagi menjadi tiga kelompok, yaitu mudah, sedang, dan sukar [34]. Dari hasil pengujian

tingkat kesukaran butir soal apabila dibagi berdasarkan masing-masing kelompok indeks tingkat kesukaran, maka diketahui seberapa besar persentase tingkat kesukaran pada masing-masing kelompok. Adapun distribusi butir soal berdasarkan indeks tingkat kesukaran ditunjukkan Tabel 4.

Tabel 4. Distribusi soal berdasarkan indeks tingkat kesukaran

No.	Indeks Tingkat Kesukaran	No. Butir Soal	Jumlah	Persentase
1.	Mudah (0,70 – 1)	1, 2, 6, 9, 13, 14, 19, 21, 24, 28, 29, 30, 31, 33, 34, 37, 42, 43, 46	19	38%
2.	Sedang (0,3 – 0,69)	4, 5, 7, 8, 10, 12, 16, 17, 18, 20, 22, 25, 27, 32, 36, 38, 40, 41, 45, 47, 48, 49, 50	23	46%
3.	Sukar (0,00 – 0,29)	3, 11, 15, 23, 26, 35, 39, 44	8	16%
Jumlah Total			50	100%

Sesuai data pada Tabel 4 di atas, apabila data disajikan dalam bentuk diagram lingkaran atau diagram *pie*, maka nampak seperti pada Gambar 2 berikut ini.



Gambar 2. Diagram *pie* distribusi soal berdasarkan indeks tingkat kesukaran

Berdasarkan data pada Tabel 4 dan Gambar 2 didapatkan indeks tingkat kesukaran yang menunjukkan bahwa dari total keseluruhan soal tipe HOTS pada kompetensi sistem rem, soal dengan kategori "Mudah" memiliki persentase 38% sejumlah 19 soal. Sedangkan soal dengan kategori "Sedang" memiliki persentase 46% sejumlah 23 soal dan soal dengan kategori "Sukar" memiliki persentase 16% sejumlah 8 soal. Hasil tersebut selaras dengan teori yang dikemukakan oleh Sudjana (2017), dimana lebih mengarah pada proporsi perbandingan 3-5-2 atau 30% butir soal dengan kategori mudah, 50% butir soal dengan kategori sedang, dan 20% butir soal dengan kategori sukar [12].

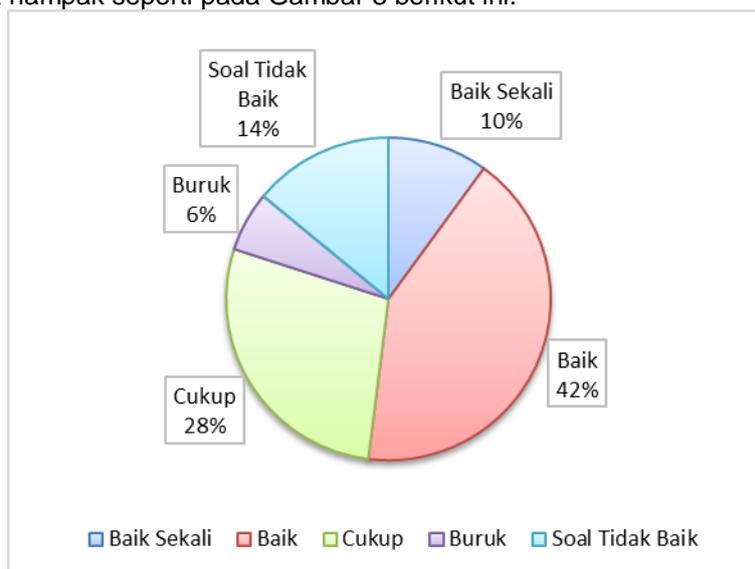
E. Daya Pembeda

Daya pembeda butir soal merupakan kemampuan dari suatu soal dalam memetakan antara siswa yang masuk pada kategori pandai hingga kurang pandai [35]. Dalam penelitian ini kemampuan daya pembeda dibagi dalam lima kategori, diantaranya: (1) baik sekali; (2) baik; (3) cukup; (4) buruk; dan (5) soal tidak baik [36]. Analisis daya pembeda butir soal dilakukan menggunakan bantuan *software Anates V4* [37]. Dari hasil pengujian daya beda butir soal apabila hasil analisis dibagi dalam lima kategori, maka diketahui seberapa besar persentase daya beda pada setiap butir soal. Adapun distribusi soal berdasarkan indeks daya beda dapat dilihat pada Tabel 5.

Tabel 5. Distribusi soal berdasarkan indeks daya beda

No.	Daya Beda	No. Butir Soal	Jumlah	Persentase
1.	Baik Sekali	8, 22, 25, 32, 38	5	10%
2.	Baik	1, 6, 10, 12, 13, 16, 17, 18, 20, 23, 24, 27, 28, 34, 36, 37, 45, 47, 48, 49, 50	21	42%
3.	Cukup	2, 9, 14, 19, 21, 29, 31, 33, 40, 41, 42, 43, 46	14	28%
4.	Buruk	3, 4, 44	3	6%
5.	Soal Tidak Baik	5, 11, 15, 26, 30, 35, 39	7	14%
Jumlah Total			50	100%

Sesuai data pada Tabel 5 di atas, apabila ditampilkan dalam bentuk diagram lingkaran atau diagram *pie*, maka nampak seperti pada Gambar 3 berikut ini.



Gambar 3. Diagram *pie* distribusi soal berdasarkan daya beda

Berdasarkan Tabel 5 diketahui bahwa dari total keseluruhan soal tipe HOTS pada kompetensi sistem rem, soal dengan kategori “Baik” memiliki kontribusi yang cukup signifikan dimana memiliki persentase sebesar 42%. Sedangkan soal dengan kategori “Cukup” memiliki persentase sebesar 28%, soal dengan kategori “Buruk” memiliki persentase sebesar 6%, dan soal dengan kategori “Tidak Baik” memiliki persentase sebesar 14%. Kemudian soal dengan kategori “Baik Sekali” memiliki persentase yang paling rendah yakni sebesar 10%. Soal dapat dikatakan telah memenuhi kriteria apabila terletak pada kategori baik, cukup, dan baik sekali. Sedangkan soal dikatakan tidak memenuhi kriteria apabila masuk kategori buruk atau soal tidak baik. Soal yang tidak memenuhi kriteria berarti soal tersebut belum mampu membedakan siswa yang memiliki kemampuan tinggi dengan siswa yang memiliki kemampuan rendah. Sidabutar, Putrayasa, & Martha (2017) melalui hasil penelitiannya menemukan bahwa setidaknya terdapat tiga faktor yang menjadi penyebab soal tidak memiliki daya pembeda yang berkualitas. *Pertama*, kualitas soal rendah yang berarti soal tersebut terlalu sulit sehingga baik siswa pada kelompok atas maupun bawah sama-sama tidak bisa menjawab dengan benar. *Kedua*, materi yang diujikan melenceng dari apa yang telah diajarkan, sehingga siswa bingung dalam menjawab soal dengan benar. *Ketiga*, kemampuan, kematangan, motivasi siswa pada kelompok bawah lebih baik dari pada kelompok atas, sehingga skor yang didapatkan kelompok bawah cenderung lebih baik bila dibandingkan dengan kelompok atas [38].

F. Efektivitas Pengecoh

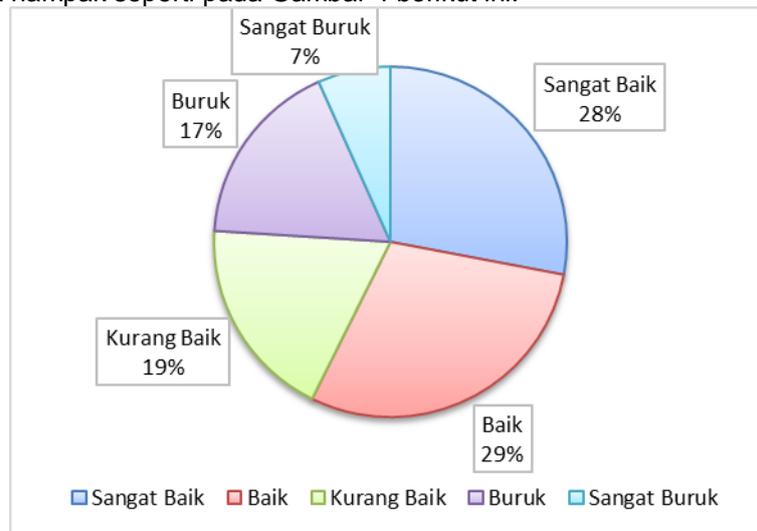
Pengecoh (distraktor) umumnya digunakan untuk mengecoh melalui jawaban yang tersedia pada soal. Analisis efektivitas pengecoh pada soal pilihan ganda tipe HOTS pada kompetensi sistem rem dilakukan menggunakan bantuan *software* Anates V4 [37]. Berdasarkan hasil analisis, apabila dibagi dalam empat kelompok indeks pengecoh maka dapat diketahui besaran persentase efektivitas

pengecoh tiap kategori. Perhitungan efektivitas pengecoh dilakukan berdasarkan jumlah pilihan jawaban yang kemudian dikalikan dengan jumlah butir soal, akan tetapi kunci jawaban tidak termasuk dalam perhitungan sehingga total jumlahnya menjadi 150 pengecoh. Adapun distribusi soal berdasarkan indeks pengecoh seperti pada Tabel 6.

Tabel 6. Distribusi soal berdasarkan indeks pengecoh

No.	Pengecoh	Jumlah	Persentase
1.	Sangat Baik	42	28%
2.	Baik	44	29%
3.	Kurang Baik	28	19%
4.	Buruk	26	17%
5.	Sangat Buruk	10	7%
Total Jumlah		150	100%

Sesuai data pada Tabel 6 di atas, apabila disajikan dalam bentuk diagram lingkaran atau diagram *pie*, maka nampak seperti pada Gambar 4 berikut ini.



Gambar 4. Diagram *pie* distribusi soal berdasarkan indeks pengecoh

Berdasarkan Tabel 6 dan Gambar 4, dapat diketahui masing-masing persentase efektivitas pengecoh pada setiap kategori. Hasil diatas menunjukkan bahwa dari 150 pengecoh yang ada pada 50 butir soal, sejumlah 42 soal atau 28% pengecoh masuk dalam kategori “Sangat Baik”. Sejumlah 44 soal atau 29% pengecoh masuk dalam kategori “Baik”. Kemudian 28 soal atau 19% pengecoh masuk dalam kategori “Kurang baik”. Sedangkan 26 soal atau 17% untuk pengecoh dengan kategori “Buruk”, dan 10 soal atau 7% pengecoh masuk dalam kategori “Sangat Buruk”. Menurut Arikunto (2013) suatu pengecoh dikatakan berfungsi apabila suatu jawaban paling sedikit dipilih oleh 5% responden [15]. Hasil ini menunjukkan bahwa rata-rata pengecoh yang tersedia pada 50 butir soal pilihan ganda tipe HOTS telah masuk kriteria efektivitas pengecoh. Namun terdapat salah satu pengecoh yang kurang bekerja secara signifikan, hal ini menunjukkan bahwa pengecoh terlalu mencolok perbedaannya bila dibandingkan pilihan jawaban yang lainnya sehingga kurang menimbulkan ketertarikan siswa untuk memilih pengecoh tersebut sehingga persebaran pengecoh hanya sebesar 7% terhadap alternatif jawaban [39], [40].

4. SIMPULAN

Berdasarkan hasil pengujian dan analisa dalam pembahasan, maka dapat ditarik kesimpulan sebagai berikut. Soal tipe HOTS pada kompetensi sistem rem layak digunakan sebagai alat penilai kompetensi siswa. Hal ini dapat dibuktikan dari total 50 soal pilihan ganda terdapat 80% soal yang masuk dalam kategori valid, sedangkan 20% soal masuk pada kategori tidak valid. Kemudian untuk reliabilitas mendapatkan nilai koefisien sebesar 0,917. Sementara itu, untuk tingkat kesukaran terdapat 38% soal masuk kategori mudah, 46% soal masuk kategori sedang, dan 16% soal masuk kategori sukar. Selanjutnya untuk daya beda 10% soal masuk kategori baik sekali, 42% soal masuk

kategori baik, 28% soal masuk kategori cukup, 6% soal masuk kategori buruk, dan 14% soal masuk kategori tidak baik. Kemudian untuk efektivitas pengecoh terdapat 150 pengecoh dari total 50 butir soal. Dari total keseluruhan terdapat 28% pengecoh masuk kategori sangat baik, 29% pengecoh masuk kategori baik, 19% pengecoh masuk kategori kurang baik, 17% pengecoh masuk kategori buruk, dan 7% pengecoh masuk kategori sangat buruk.

Berdasarkan hasil penelitian dan simpulan yang disampaikan, peneliti memberikan saran sebagai berikut: (1) sebelum memberikan tes atau ujian kepada siswa, guru wajib untuk melakukan analisis butir terlebih dahulu agar instrumen yang digunakan dapat diketahui tingkat kevalidannya; (2) guru diarahkan agar mampu meningkatkan pemahaman dan kemampuan dalam penyusunan instrumen tes agar butir-butir soal yang dihasilkan lebih berkualitas; dan (3) pelaksanaan tes harus bersifat adil bagi seluruh siswa dengan proporsi perbandingan soal yang seimbang sesuai apa yang telah dipelajari.

DAFTAR PUSTAKA

- [1] I. Maulana, R. Arthur, and Daryati, "Kualitas Tes Kompetensi Tukang Aci dan Plesteran Berdasarkan KKNi," *J. Pendidik. Teknol. dan Kejuru.*, vol. 16, no. 2, pp. 149–158, 2019.
- [2] A. Surya and A. Aman, "Developing formative authentic assessment instruments based on learning trajectory for elementary school," *Res. Eval. Educ.*, vol. 2, no. 1, p. 13, Jun. 2016.
- [3] F. Tangianu *et al.*, "Are multiple-choice questions a good tool for the assessment of clinical competence in Internal Medicine?," *Ital. J. Med.*, vol. 12, no. 2, p. 88, Jun. 2018.
- [4] G. T. L. Brown and H. H. A. Abdulnabi, "Evaluating the Quality of Higher Education Instructor-Constructed Multiple-Choice Tests: Impact on Student Grades," *Front. Educ.*, vol. 2, Jun. 2017.
- [5] L. W. Anderson and D. R. Krathwohl, *A taxonomy for learning teaching and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman, Inc, 2001.
- [6] A. Javaeed, "Assessment of Higher Ordered Thinking in Medical Education: Multiple Choice Questions and Modified Essay Questions," *MedEdPublish*, vol. 7, no. 2, 2018.
- [7] A. Friatma and A. Anhar, "Analysis of validity, reliability, discrimination, difficulty and distraction effectiveness in learning assessment," *J. Phys. Conf. Ser.*, vol. 1387, p. 012063, Nov. 2019.
- [8] F. N. Kerlinger, *Foundations of behavioral research*, 2nd ed. New York: Holt, Rinehart and Winston, Inc, 1973.
- [9] M. Nazir, *Metode penelitian*, 11th ed. Bogor: Ghalia Indonesia, 2014.
- [10] K. S. Taber, "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education," *Res. Sci. Educ.*, vol. 48, no. 6, pp. 1273–1296, Dec. 2018.
- [11] S. Arikunto, *Prosedur penelitian suatu pendekatan praktik*, 7th ed. Jakarta: Rineka Cipta, 2014.
- [12] N. Sudjana, *Penilaian hasil proses belajar mengajar*, 20th ed. Bandung: Remaja Rosdakarya, 2017.
- [13] C. Rao, H. Kishan Prasad, K. Sajitha, H. Permi, and J. Shetty, "Item analysis of multiple choice questions: Assessing an assessment tool in medical students," *Int. J. Educ. Psychol. Res.*, vol. 2, no. 4, p. 201, 2016.
- [14] Z. Arifin, *Evaluasi pembelajaran*, 2nd ed. Jakarta: Direktorat Jenderal Pendidikan Islam Kementerian Agama RI, 2012.
- [15] S. Arikunto, *Dasar-dasar evaluasi pendidikan*, 2nd ed. Jakarta: Bumi Aksara, 2013.
- [16] Purwanto, *Evaluasi hasil belajar*, 6th ed. Yogyakarta: Pustaka Pelajar, 2014.
- [17] D. Kheyami, A. Jaradat, T. Al-Shibani, and F. A. Ali, "Item Analysis of Multiple Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain.," *Sultan Qaboos Univ. Med. J.*, vol. 18, no. 1, pp. e68–e74, Feb. 2018.
- [18] Y. Karkal and G. Kundapur, "Item analysis of multiple choice questions of undergraduate pharmacology examinations in an International Medical School in India," *J. Dr. NTR Univ. Heal. Sci.*, vol. 5, no. 3, p. 183, 2016.
- [19] J. A. Odukoya, O. Adekeye, A. O. Igbinoba, and A. Afolabi, "Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university," *Qual. Quant.*, vol. 52, no. 3, pp. 983–997, May 2018.
- [20] A. Sumiati, U. Widiastuti, and U. Suhud, "Workshop Teknik Menganalisis Butir Soal dalam Meningkatkan Kompetensi Guru di SMK Cileungsi Bogor," *J. Pemberdaya. Masy. Madani*, vol. 2, no. 1, pp. 136–153, Jul. 2018.
- [21] R. Y. Kurniawan, A. Fiky Prakoso, L. Hakim, R. Mustika Dewi, and I. Widayanti, "Pemberian Pelatihan Analisis Butir Soal Bagi Guru di Kabupaten Jombang : Efektif?," *J. Pemberdaya. Masy. Madani*, vol. 1, no. 2, pp. 179–193, Dec. 2017.
- [22] M. Kaur, S. Singla, and R. Mahajan, "Item analysis of in use multiple choice questions in

- pharmacology.," *Int. J. Appl. basic Med. Res.*, vol. 6, no. 3, pp. 170–3, 2016.
- [23] K. Quaigrain and A. K. Arhin, "Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation," *Cogent Educ.*, vol. 4, no. 1, pp. 1–11, 2017.
- [24] A. A. Danuwijaya, "Item Analysis of Reading Comprehension Test for Post-Graduate Students," *English Rev. J. English Educ.*, vol. 7, no. 1, p. 29, Dec. 2018.
- [25] M. Kusumawati and S. Hadi, "An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school," *Res. Eval. Educ.*, vol. 4, no. 1, pp. 70–78, Jul. 2018.
- [26] N. Sa'idah, H. D. Yulistianti, and E. Megawati, "Analisis Instrumen Tes Higher Order Thinking Matematika SMP," *J. Pendidik. Mat.*, vol. 13, no. 1, pp. 41–54, 2018.
- [27] H. Taherdoost, "Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research," *SSRN Electron. J.*, 2016.
- [28] M. Naqiyah, Jumadi, and I. Wilujeng, "Physics Learning Using Direct Instruction Model Assisted by Plickers Application to Measure Problem Solving Ability," *J. Phys. Conf. Ser.*, vol. 1227, p. 012031, Jun. 2019.
- [29] A. C. de Souza, N. M. C. Alexandre, E. de B. Guirardello, A. C. de Souza, N. M. C. Alexandre, and E. de B. Guirardello, "Psychometric properties in instruments evaluation of reliability and validity.," *Epidemiol. e Serviços Saúde*, vol. 26, no. 3, pp. 649–659, Jul. 2017.
- [30] R. Setiawaty, T. B. Sulistyorini, Margono, and L. . Rahmawati, "Validity Test and Reliability of Indonesian Language Multiple Choice in Final Term Examination," *KnE Soc. Sci.*, vol. 3, no. 9, p. 43, 2018.
- [31] O. Bolarinwa, "Principles and methods of validity and reliability testing of questionnaires used in social and health science researches," *Niger. Postgrad. Med. J.*, vol. 22, no. 4, p. 195, 2015.
- [32] A. M. Al-Osail *et al.*, "Is Cronbach's alpha sufficient for assessing the reliability of the OSCE for an internal medicine course?," *BMC Res. Notes*, vol. 8, p. 582, Oct. 2015.
- [33] D. P. Sahoo and R. Singh, "Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students," *Int. J. Res. Med. Sci.*, vol. 5, no. 12, p. 5351, Nov. 2017.
- [34] I. Pistol, D. Trandabăţ, and M. Răschip, "Medi-Test: Generating Tests from Medical Reference Texts," *Data*, vol. 3, no. 4, p. 70, Dec. 2018.
- [35] P. Mukherjee and S. K. Lahiri, "Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal," *IOSR J. Dent. Med. Sci.*, vol. 14, no. 12, pp. 2279–861, 2015.
- [36] W. Mahjabeen *et al.*, "Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions," *Ann. Pakistan Inst. Med. Sci.*, 2017.
- [37] E. Yusmaita and E. Nasra, "Design of Chemical Literacy Assessment by Using Model of Educational Reconstruction (MER) on Solubility Topic," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 335, p. 012106, Apr. 2018.
- [38] G. D. U. Sidabutar, I. B. Putrayasa, and I. N. Martha, "Kualitas Butir Soal Ulangan Akhir Semester Ganjil Bahasa Indonesia Kelas IX SMP Negeri 2 Singaraja Tahun Pelajaran 2016 / 2017 Ditinjau Dari Segi Taraf Kesukaran, Daya Beda, Dan Fungsi Pengecoh," *e-Journal Jur. Pendidik. Bhs. dan Sastra Indones. Undiksha*, vol. 7, no. 2, 2017.
- [39] M. Purwanti, "Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan Microsoft Office Excel 2010," *J. Pendidik. Akunt. Indones.*, vol. 12, no. 2, 2014.
- [40] A. Iskandar and M. Rizal, "Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP," *J. Penelit. dan Eval. Pendidik.*, vol. 22, no. 1, p. 12, 2018.