

ANALISIS SOAL PILIHAN GANDA UJIAN AKHIR SEMESTER GENAP MATA PELAJARAN MEKANIKA TEKNIK KELAS X JURUSAN TEKNOLOGI KONSTRUKSI DAN PROPERTI DI SMK NEGERI 1 MAGELANG TAHUN AJARAN 2017/2018

ANALYSIS ABOUT MULTIPLE CHOICE A FINAL EXAM THE FIRST HALF THE EVEN SUBJECTS MECHANICS TECHNIQUE CLASS X OF TECHNOLOGY CONSTRUCTION AND PROPERTY IN SMK NEGERI 1 MAGELANG ACADEMIC YEAR 2017/2018

Oleh: Deny Astuti

Pendidikan Teknik Sipil dan Perencanaan, Fakultas Teknik, Universitas Negeri Yogyakarta

email: 87denyastuti@gmail.com

Abstrak

Tujuan penelitian untuk mengetahui kualitas soal pilihan ganda Ujian Akhir Semester (UAS) mata pelajaran mekanika teknik. Jenis penelitian ini adalah deskriptif kuantitatif. Penelitian ini menggunakan teknik dokumentasi. Penelitian dilakukan dengan analisis instrumen dan analisis empiris. Analisis instrumen ditinjau dari penelaahan soal, dan pendistribusian tingkat berikir ranah kognitif. Analisis empiris ditinjau dari tingkat kesukaran, daya pembeda, efektivitas pengecoh, validitas butir, dan reliabilitas soal. Hasil penelitian menunjukkan bahwa butir soal memenuhi aspek materi sebesar 86,67%, aspek konstruksi sebesar 56,67%, dan aspek bahasa sebesar 86,67%. Distribusi tingkat berpikir ranah kognitif kategori C1 33,33%, C2 20%, C3 46,67%, C4 0%, C5 0%, dan C6 0%. Tingkat kesukaran soal kategori mudah 40%, sedang 33,33%, dan sukar 26,67%. Daya pembeda soal kategori baik 26,67%, cukup 36,67%, kurang baik 33,33%, dan tidak baik 3,33%. Efektivitas pengecoh sebesar 66,67% berfungsi baik, dan 33,33% tidak berfungsi baik. Validitas butir sebesar 66,67% soal dinyatakan valid, dan 33,33% tidak valid. Reliabilitas soal kategori kurang reliabel dengan koefisien korelasi sebesar 0,621.

Kata kunci: analisis soal, UAS, mekanika teknik

Abstract

The research objective is to find out the quality of multiple choice questions in the Semester End Exams (UAS) in mechanical engineering subjects. This type of research is quantitative descriptive. This study uses documentation techniques. The study was conducted with instrument analysis and empirical analysis. Analysis of the instrument in terms of the review of the problem, and the distribution of thinking levels in the cognitive domain. Empirical analysis in terms of the level of difficulty, distinguishing power, deception effectiveness, item validity, and reliability of the problem. The results showed that the items met the material aspects of 86.67%, the construction aspect was 56.67%, and the language aspect was 86.67%. The distribution of cognitive level thinking in the C1 category was 33.33%, C2 20%, C3 46.67%, C4 0%, C5 0%, and C6 0%. The level of difficulty in easy category questions is 40%, moderate is 33.33%, and it is difficult for 26.67%. Differentiation of good category 26.67%, enough 36.67%, poor 33.33%, and not good 3.33%. Deceptive effectiveness of 66.67% functions well, and 33.33% does not function well. Item validity of 66.67% is declared valid, and 33.33% is invalid. Reliability of category questions is less reliable with a correlation coefficient of 0.621.

Keywords: problem analysis, UAS, technical mechanics

PENDAHULUAN

Sekolah menengah kejuruan (SMK) merupakan sekolah yang mempersiapkan siswa menjadi lulusan yang siap terjun ke dunia kerja sebagai individu yang produktif, kreatif, dan memiliki keahlian khusus sehingga mampu bersaing dalam dunia kerja. Meskipun lebih dipersiapkan untuk terjun ke dunia kerja, lulusan

SMK juga dapat melanjutkan belajar ke perguruan tinggi sesuai keahliannya.

Terdapat tiga bagian penting dalam dunia pendidikan, yaitu kurikulum, proses pembelajaran, dan evaluasi hasil belajar. Kurikulum sebagai perencana tujuan pendidikan, proses pembelajaran sebagai pelaksanaan tujuan

pendidikan, dan evaluasi hasil belajar sebagai penilai dan kontrol terhadap tujuan pendidikan.

Menurut PP RI No. 19 Tahun 2017 tentang Perubahan Atas Peraturan Pemerintah Nomor 74 Tahun 2008 tentang Guru pada Pasal 1 ayat 1 menyatakan bahwa “Guru adalah pendidik profesional dengan tugas utama mendidik, mengajar, membimbing, mengarahkan, melatih, menilai, dan mengevaluasi siswa pada pendidikan anak usia dini jalur pendidikan formal, pendidikan dasar, dan pendidikan menengah”. Oleh karena itu, seharusnya sebagai guru yang paling mengerti seluruh kegiatan pembelajaran mampu membuat dan menyusun soal yang berkualitas agar mampu mewakili hasil ketercapaian kompetensi yang harus diukur dan dinilai dari proses pembelajaran tersebut.

Berdasarkan hasil tanya-jawab dengan salah satu guru yang mengampu mata pelajaran mekanika teknik Jurusan Teknologi Konstruksi dan Properti SMK Negeri 1 Magelang pada tanggal 8 April 2018, diketahui bahwa proses penilaian mata pelajaran mekanika teknik pada UAS Genap Tahun Ajaran 2017/2018 menggunakan tes soal yang dibuat oleh guru mata pelajaran mekanika teknik itu sendiri. Soal tes tersebut disusun dan disiapkan tanpa melalui kajian yang lebih rinci dan terstruktur seperti melakukan analisis soal tes terlebih dahulu. Sehingga soal UAS genap pada mata pelajaran mekanika teknik Jurusan Teknologi Konstruksi dan Properti tersebut belum diketahui kualitasnya.

Soal UAS genap ini memiliki peranan penting untuk mengukur ketercapaian siswa akan kompetensi yang diajarkan, terlebih lagi sebagai salah satu penentuan kenaikan kelas bagi siswa. Beberapa faktor kemungkinan penyebab guru belum melaksanakan analisis soal UAS adalah keterbatasan waktu dan biaya dalam menganalisis soal UAS, kurangnya pengalaman dan pengetahuan mengenai analisis soal.

Menurut Tegar Putra Siregar, dkk (2017: 2153) menyimpulkan bahwa analisis soal merupakan suatu proses yang dilakukan untuk menyelidiki, meneliti, dan meninjau soal tes untuk mendapatkan perangkat tes yang

berkualitas memadai. Menurut Budi Susetyo (2015: 179) tujuan dilaksanakannya analisis butir soal adalah untuk mencari butir tes yang berkualitas dan digunakan sebagai alat ukur hasil belajar atau instrumen penelitian dalam bidang kognitif. Disamping itu, analisis butir soal juga digunakan untuk melihat butir mana yang menyebabkan reliabilitas pengukuran menjadi rendah. Dilaksanakannya telaah dan uji coba soal, merupakan bagian dari analisis butir soal yang kemudian digunakan untuk melakukan perbaikan, penambahan, pembuangan, dan penggantian butir sehingga alat ukur menjadi lebih berkualitas.

Kusaeri & Suprananto (2012: 163) menjelaskan bahwa analisis butir terdiri atas tingkat kesukaran butir, daya pembeda, dan distraktor. Aspek materi, konstruksi, dan bahasa termasuk dalam kategori analisis penelaahan soal. Ketiga aspek tersebut disajikan dalam satu format lembar penelaahan soal. Selain ketiga aspek tersebut, butir soal juga diukur dari sisi tingkat berpikir yang diperlukan peserta tes untuk menjawab soal. Hal ini dilakukan dengan mendistribusikan butir soal ke dalam salah satu dari enam jenjang ranah kognitif taksonomi Bloom. Semakin butir soal membutuhkan tingkat berpikir yang tinggi maka soalpun berkualitas tinggi.

METODE PENELITIAN

Metode pada penelitian ini menggunakan metode kuantitatif. Dalam bab ini akan membahas tentang jenis penelitian, waktu penelitian, subjek dan objek penelitian, teknik pengumpulan data dan teknik analisis data.

Jenis Penelitian

Penelitian ini termasuk dalam jenis penelitian deskriptif kuantitatif. Menurut pendapat Suharsimi Arikunto (2013: 3) penelitian deskriptif adalah penelitian yang bermaksud untuk menyelidiki keadaan, kondisi atau hal-hal lain yang sudah terjadi, kemudian hasilnya ditulis dalam bentuk laporan penelitian.

Analisis yang dilakukan pada penelitian ini adalah analisis soal pilihan ganda UAS genap mata pelajaran mekanika teknik kelas X Jurusan

Teknologi Konstruksi dan Properti di SMK Negeri 1 Magelang tahun ajaran 2017/2018. Pada analisis instrumen, dilakukan analisis penelaahan soal berdasarkan kaidah penulisan soal meliputi aspek materi, konstruksi, dan bahasa menggunakan lembar telaah soal pilihan ganda yang dilakukan oleh peneliti, serta mengidentifikasi pendistribusian tingkat berpikir ranah kognitif menggunakan taksonomi Bloom revisi pada instrumen. Kemudian dilakukan analisis empiris meliputi tingkat kesukaran, daya pembeda, efektivitas pengecoh, validitas butir, dan reliabilitas soal dengan bantuan program aplikasi ANATES versi 4.0.9. Peneliti menggunakan program aplikasi ANATES versi 4.0.9 karena lebih mudah digunakan dan mampu menganalisis soal-soal pilihan ganda.

Tujuan analisis butir soal UAS ini adalah untuk mengetahui kualitas soal UAS genap mata pelajaran mekanika teknik kelas X yang dibuat oleh guru pengampu mata pelajaran mekanika teknik itu sendiri.

Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan di SMK Negeri 1 Magelang yang beralamat di Jalan Cawang No 2, Kelurahan Jurang Ombo, Kecamatan Magelang Selatan, Jawa Tengah pada bulan April – Juli 2018 sebelum dan setelah pelaksanaan UAS genap kelas X Jurusan Teknologi Konstruksi dan Properti tahun ajaran 2017/2018.

Subjek dan Objek Penelitian

Subjek pada penelitian ini adalah seluruh siswa kelas X Jurusan Teknologi Konstruksi dan Properti SMK Negeri 1 Magelang tahun ajaran 2017/2018. Terdapat tiga kelas X pada Jurusan Teknologi Konstruksi dan Properti yaitu kelas X BA, X BB, dan X BC. Siswa di ketiga kelas tersebut berjumlah 94 siswa. Sedangkan objek dalam penelitian ini adalah soal pilihan ganda UAS genap mata pelajaran mekanika teknik kelas X Jurusan Teknologi Konstruksi dan Properti SMK Negeri 1 Magelang tahun ajaran 2017/2018. Soal pilihan ganda mekanika teknik ini berjumlah 30 butir soal. Masing-masing soal memiliki 5 jawaban opsional dengan satu jawaban benar.

Teknik Pengumpulan Data

Salah satu teknik pengumpulan data yang termasuk dalam data sekunder adalah teknik dokumentasi. Data sekunder atau data tangan pertama adalah data yang diperoleh peneliti. Dokumentasi yang diperoleh berupa silabus mekanika teknik, kisi-kisi soal UAS Genap mekanika teknik kelas X tahun ajaran 2017/2018, soal UAS, kunci jawaban, dan lembar jawab siswa.

Teknik Analisis Data

Teknik analisis data pada penelitian ini menggunakan metode kuantitatif. Analisis ini dibagi menjadi dua bagian, yaitu analisis instrumen, dan analisis empiris.

1. Analisis Instrumen

Pada analisis instrumen dibutuhkan format lembar penelaahan butir soal dari aspek materi, konstruksi, dan bahasa yang dikutip dari Kusaeri & Suprananto (2012: 168-169) dan telah dikembangkan oleh peneliti. Selain itu juga dibutuhkan format lembar pendistribusian tingkat berpikir ranah kognitif sesuai taksonomi Bloom revisi berupa tabel yang dibuat oleh peneliti.

2. Analisis Empiris

Pada analisis empiris, dilakukan perhitungan menggunakan bantuan program aplikasi ANATES versi 4.0.9 serta perhitungan manual menggunakan bantuan program aplikasi Microsoft Excel.

a. Tingkat Kesukaran

Tingkat kesukaran pada butir soal dapat di kategorikan dalam tiga kriteria, yaitu butir soal kategori mudah, kategori sedang, dan kategori sukar. Rumus yang digunakan untuk mencari indeks kesukaran yang diberi simbol P (Proporsi) dikutip dari Suharsimi Arikunto (2016: 223), yaitu:

$$P = \frac{B}{JS} \quad (1)$$

dengan P = indeks kesukaran, B = banyaknya siswa yang menjawab benar soal, dan JS = jumlah seluruh siswa peserta tes.

Tabel 1. Klasifikasi Indeks Kesukaran

Indeks Kesukaran	Kategori	Keputusan
0,70 - 1,00	Mudah	Ditolak/direvisi
0,30 - 0,70	Sedang	Diterima
0,00 - 0,30	Sukar	Ditolak/direvisi

b. Daya Pembeda

Daya pembeda dihitung guna mengetahui perbedaan hasil tes pada siswa kelompok atas dan siswa kelompok bawah. Butir soal yang dapat dijawab dengan benar oleh kedua kelompok seharusnya merupakan butir soal mudah. Rumus untuk mencari indeks diskriminasi daya pembeda (D) yang dikutip dari Suharsimi Arikunto, (2016: 232), yaitu:

$$D = \frac{Ba}{Ja} - \frac{Bb}{Jb} = Pa - Pb \quad (2)$$

dengan D = indeks diskriminasi daya pembeda, Ja = Banyaknya peserta tes kelompok atas, Jb = Banyaknya peserta tes kelompok bawah, Ba = Banyaknya peserta kelompok atas yang menjawab benar, Bb = Banyaknya peserta kelompok bawah yang menjawab benar, Pa = Kelompok atas yang menjawab benar, Pb = Kelompok bawah yang menjawab benar.

Tabel 2. Klasifikasi Daya Pembeda

Indeks Daya Beda (D)	Kriteria
D = negatif	Semua tidak baik (dibuang)
0,00 - 0,20	Tidak Baik (poor)
0,20 - 0,40	Cukup (satisfactory)
0,40 - 0,70	Baik (Good)
0,70 - 1,00	Baik Sekali (excellent)

c. Efektivitas Pengecoh

Efektivitas pengecoh dikatakan berfungsi atau tidak dapat ditandai dengan mengetahui hasil persentase dari indeks pengecoh. Perhitungan efektivitas pengecoh menggunakan bantuan

program aplikasi ANATES versi 4.0.9. Hasil data yang diperoleh dari aplikasi tersebut kemudian dapat diinterpretasikan menggunakan kriteria yang diadaptasi dari Skala Likert yang dikutip dari Sugiyono (2015: 134-135) yang diterapkan pada Tabel 12. Hal ini diketahui dari jumlah jawaban pengecoh pada butir soal dalam bentuk persentasenya. Pilihan jawaban dianggap berfungsi sebagai pengecoh apabila dipilih $\geq 5\%$ dari seluruh peserta tes.

Tabel 3. Klasifikasi Hasil Penilaian Efektivitas Pengecoh

Kategori	Pengecoh yang Berfungsi Tiap Butir Soal
Sangat baik	4
Baik	3
Cukup Baik	2
Kurang Baik	1
Tidak Baik	0

d. Validitas butir

Hasil perhitungan analisis validitas butir digunakan untuk mengetahui apakah butir soal tes valid atau tidak. Cara penyelesaiannya adalah dengan mengorelasikan antara skor butir (X) dengan skor total (Y). Kemudian nilai koefisien korelasi yang diperoleh tersebut dicocokkan dengan nilai r tabel product moment dengan tingkat signifikansi 5%. Butir soal dapat dinyatakan valid apabila diperoleh nilai r tabel $< r$ hitung. Sebaliknya, jika nilai r hitung $< r$ tabel maka dapat butir soal dinyatakan tidak valid. Rumus untuk mengetahui validitas butir soal adalah dengan menggunakan teknik korelasi product moment yang dikemukakan oleh Person dikutip dari Lili Maenani & Raden Oktavia (2015: 7-8):

$$r_{XY} = \frac{n (\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{[n (\Sigma X^2) - (\Sigma X)^2] [n (\Sigma Y^2) - (\Sigma Y)^2]}} \quad (3)$$

dengan r_{xy} = koefisien korelasi antara variabel X dan variabel Y , n = jumlah responden, Y = skor total yang diperoleh responden dari semua butir

soal, X = skor butir soal yang diperoleh responden.

e. Reliabilitas

Dalam mencari koefisien reliabilitas soal tes secara keseluruhan maka dapat menggunakan rumus KR-20. Rumus KR-20 digunakan mengetahui konsistensi hasil pengukuran dari suatu instrumen soal yang dikemukakan oleh Kuder Richardson-20 (KR-20) dikutip dari Lili Maenani & Raden Oktavia (2015: 7-8) sebagai berikut.

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \Sigma pq}{S^2} \right) \quad (4)$$

dengan r_{11} = koefisien reliabilitas tes secara keseluruhan, S = standar deviasi, p = proporsi subjek yang menjawab benar, q = proporsi subjek yang menjawab salah, Σpq = Jumlah hasil perkalian antara p dan q .

Hasil perhitungan koefisien reliabilitas kemudian diinterpretasikan sesuai dengan kriteria yang dikutip dari Anas Sudijono (2015: 209), bahwa apabila $r_{11} \geq 0,70$ berarti tes hasil belajar yang diuji reliabilitasnya telah memiliki reliabilitas yang tinggi (= *reliable*), dan apabila $r_{11} \leq 0,70$ berarti tes hasil belajar belum memiliki reliabilitas yang tinggi (*un-reliable*).

HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini dilaksanakan dengan tujuan mengetahui kualitas alat evaluasi hasil belajar. Alat evaluasi hasil belajar yang digunakan dalam penelitian ini adalah soal tes ujian akhir semester genap mata pelajaran mekanika teknik kelas X tahun ajaran 2017/2018 di Jurusan Teknologi Konstruksi dan Properti SMK Negeri 1 Magelang.

Kualitas butir soal ujian akhir semester genap mata pelajaran mekanika teknik ini difokuskan pada soal objektif. Analisis dilihat dari aspek materi, aspek konstruksi, dan aspek bahasa serta dilihat dari pendistribusian tingkat berpikir ranah kognitif pada analisis instrumen, dan menyangkut tingkat kesukaran, daya pembeda, efektivitas pengecoh/distractor, validitas butir, dan reliabilitas soal pada analisis empiris.

Analisis Instrumen

a. Penelaahan soal

Hasil penelaahan soal menunjukkan pada aspek materi, butir soal mekanika teknik yang ditelaah terdapat 4 butir soal (13,3%) yang tidak memenuhi syarat, yaitu butir soal nomor 6, 8, 25, dan 26. Ini berarti sebanyak 26 butir soal (86,7%) sudah memenuhi seluruh aspek materi. Butir soal nomor 25 dan 26 tidak memenuhi kisi-kisi soal dan KD yang seharusnya sehingga soal harus dibuang dan diganti soal yang sesuai dengan kisi-kisi soal dan KD.

Dilihat dari aspek konstruksi diketahui bahwa ada 13 butir soal (43,3%) yang tidak sesuai, yakni butir soal nomor 3, 6, 7, 8, 11, 15, 19, 20, 23, 24, 25, 29, dan 30. Ini berarti sebanyak 17 butir soal (56,67%) sudah memenuhi seluruh aspek konstruksi. Hal ini disebabkan pada soal nomor 6, 8, 24, 25 tidak memiliki pokok soal yang jelas, kebanyakan soal sudah dirumuskan secara singkat tetapi hal tersebut justru membuat makna dari pernyataan kurang lengkap dan ambigu.

Butir soal sudah menggunakan bahasa sesuai kaidah bahasa Indonesia yang benar dan tidak menggunakan bahasa daerah setempat maupun bahasa yang tabu, yaitu ada sebanyak 26 butir soal (86,67%). Pilihan jawaban butir soal sudah sesuai kriteria karena tidak mengulang kelompok kata yang sama, hanya saja 4 butir soal (13,33%) masih menggunakan bahasa yang kurang komunikatif, karena kalimat atau pernyataan butir soal tidak lengkap dan dapat bermakna ganda atau ambigu, sehingga terkadang kunci jawaban dapat lebih dari satu.

b. Pendistribusian Tingkat Berpikir Ranah Kognitif

Pendistribusian tingkat berpikir ranah kognitif menggunakan taksonomi Bloom revisi terbagi menjadi enam jenjang berpikir. Analisis ini dilakukan untuk mencocokkan atau mengelompokkan butir-butir soal dengan kriteria tingkat berpikir ranah kognitif pada taksonomi Bloom revisi. Enam jenjang berpikir tersebut diurutkan dari kemampuan berpikir tingkat rendah (C1) sampai tingkat tinggi (C6).

Tabel 4. Hasil Pendistribusian Tingkat Berpikir

Jenjang	Nomor Soal	(%)
Mengingat (C1)	3, 6, 7, 8, 10, 11, 12, 13, 18, 20	33.33
Memahami (C2)	1, 2, 4, 5, 9, 26	20.00
Menerapkan (C3)	14, 15, 16, 17, 19, 21, 22, 23, 24, 25, 27, 28, 29, 30	46.67
Menganalisis (C4)	0	0
Mengevaluasi (C5)	0	0
Menciptakan (C6)	0	0

Berdasarkan ketiga tingkat berpikir ranah kognitif C1, C2, dan C3 yang terdapat pada soal, menurut klasifikasi pada taksonomi Bloom revisi masih tergolong tingkat berpikir rendah (*Low Order Thinking Skill*). Karena pada butir soal yang ditelaah belum menerapkan butir soal yang mengharuskan peserta tes untuk berpikir kritis atau dengan kata lain melibatkan siswa agar mampu menganalisis (C4), evaluasi (C5), dan mencipta (C6). Ketiga jenjang berpikir kritis ranah kognitif tersebut pada taksonomi Bloom revisi dikategorikan sebagai tingkat berpikir tinggi (*High Order Thinking Skill*).

Analisis Empiris

a. Tingkat Kesukaran

Hasil analisis tingkat kesukaran dapat menunjukkan kualitas butir soal pilihan ganda yang diklasifikasikan ke dalam beberapa kategori butir soal, yakni butir soal kategori mudah, sedang, dan sukar. Berikut ini merupakan hasil analisis tingkat kesukaran yang akan dipaparkan pada Tabel 5.

Tabel 5. Klasifikasi Tingkat Kesukaran Butir Soal

Indeks Kesukaran	Nomor Butir Soal	(%)	Keputusan
(Mudah) 0,70 - 1,00	1, 3, 4, 5, 9, 10, 11, 15, 16, 17, 29, 30	40	Ditolak/ direvisi
(Sedang) 0,30 - 0,70	6, 7, 12, 13, 14, 18, 22, 23, 26, 27	33.33	Diterima
(Sukar) 0,00 - 0,30	2, 8, 19, 20, 21, 24, 25, 28	26.67	Ditolak/ direvisi

Berdasarkan hasil analisis menunjukkan bahwa butir soal belum memiliki tingkat kesukaran yang ideal atau proporsional. Hal ini disebabkan butir soal dengan kategori mudah lebih banyak dibandingkan dengan butir soal kategori sedang. Selain itu, butir soal sukar dan mudah memiliki perbandingan 2:3. Sesuai dengan pernyataan Zainal Arifin (2013: 270-271) bahwa proporsi soal ideal dapat diatur seperti: soal sukar 25%, soal sedang 50%, dan soal sukar 25%. Sedangkan hasil analisis menunjukkan tingkat kesukaran pada butir soal sebanyak 40% kategori mudah, 33,33% kategori sedang, dan 26,67% kategori sukar. Sehingga sebagian butir soal yang mudah perlu direvisi dan ditingkatkan menjadi butir soal kategori sedang.

b. Daya Pembeda

Perhitungan daya pembeda dilakukan untuk mengetahui kualitas butir soal dari perbedaan skor peserta tes pada kelompok atas (siswa berkemampuan tinggi) dan kelompok bawah (siswa berkemampuan rendah). Sesuai dengan pernyataan Suharsimi Arikunto (2016: 232) bahwa butir soal yang baik memiliki indeks diskriminasi daya pembeda (D) mulai dari 0,40 sampai 0,70. Apabila nilai D dari suatu butir soal lebih dari 0,70 maka daya pembeda yang dimiliki butir soal tersebut semakin baik.

Tabel 6. Klasifikasi Daya Pembeda Butir Soal

Klasifikasi Daya Pembeda	Nomor Butir Soal	(%)
0,70 - 1,00 (Baik Sekali)	0	0
0,40 - 0,70 (Baik)	7, 9, 12, 14, 24, 26, 27, 29	26.67
0,20 - 0,40 (Cukup)	1, 3, 6, 8, 11, 19, 21, 23, 25, 28, 30	36.67
0,00 - 0,20 (Kurang Baik)	2, 4, 5, 10, 15, 16, 17, 18, 20, 22	33.33
D = negatif (Tidak Baik)	13	3.33

Dari uraian tersebut dapat disimpulkan bahwa sebanyak 19 butir soal atau 63,33% dapat dinyatakan layak untuk diujikan kepada peserta tes, dan sisanya sebesar 11 butir soal atau 36,67% dinyatakan tidak layak dan butir soal perlu direvisi. Satu dari butir soal yang tidak layak

harus dibuang karena memiliki nilai D negatif. Apabila butir soal yang tidak layak tetap digunakan, maka skor dari butir soal tersebut tidak dapat dibedakan antara siswa yang menguasai materi atau siswa yang tidak menguasai materi. Salah satu faktor yang membuat butir soal memiliki daya pembeda rendah yaitu soal terlalu sukar atau terlalu mudah.

c. Efektivitas Pengecoh

Efektivitas pengecoh berguna mengetahui kemampuan siswa sebagai peserta tes yang sebenarnya dengan jalan memberikan alternatif jawaban yang logis dan sesuai dengan materi soal, sehingga pilihan tersebut menjadi pengecoh siswa kelompok bawah atau siswa yang kurang menguasai materi kompetensi yang diukur.

Sebagai alternatif jawaban pengecoh yang baik, seharusnya minimal siswa memilih jawaban pengecoh sebanyak 5% dari seluruh peserta tes maka pengecoh yang digunakan dapat dikatakan berfungsi. Semakin banyak siswa memilih jawaban pengecoh maka alternatif jawaban tersebut berfungsi sangat baik sebagai jawaban pengecoh. Apabila kurang dari 5% siswa yang menjawab alternatif jawaban yang disediakan sebaiknya pilihan jawaban atau pengecoh tersebut perlu diganti dengan pilihan jawaban yang lebih logis.

Tabel 7. Hasil Analisis Berdasarkan Efektivitas Pengecoh

Kategori	Butir Soal	(%)
Sangat Baik	2, 28	6,67
Baik	8, 12, 13, 14, 19, 20, 24, 26, 27, 29	33,33
Cukup Baik	6, 7, 9, 10, 18, 21, 23, 25	26,67
Kurang Baik	1, 3, 5, 11, 22	16,67
Tidak Baik	4, 15, 16, 17, 30	16,67

d. Validitas Butir

Hasil perhitungan validitas menggunakan rumus korelasi dari Person menunjukkan hasil yang sama dengan perhitungan melalui program aplikasi ANATES versi 4.0.9. Berikut ini contoh perhitungan manual menggunakan bantuan

program aplikasi Microsoft Excel untuk mencari validitas pada butir soal nomor 29.

$$r_{XY} = \frac{n (\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{[n (\Sigma X^2) - (\Sigma X)^2] [n (\Sigma Y^2) - (\Sigma Y)^2]}}$$

$$r_{XY} = \frac{94 (1189) - (68)(1562)}{\sqrt{[94 (68) - (68)^2] [94 (26840) - (1562)^2]}}$$

$$r_{XY} = 0,458$$

Setelah dilakukan perhitungan validitas pada 30 butir soal, dapat diketahui jumlah butir soal yang valid sebanyak 20 butir soal (66,67%). Dari 30 butir soal, sebanyak 10 butir soal (33,33%) tidak valid. Artinya butir soal yang tidak valid tersebut perlu direvisi atau dibuang. Hasil perhitungan manual sama persis dengan rumus yang digunakan pada ANATES versi 4.0.9.

e. Reliabilitas Soal

Hasil analisis butir soal berdasarkan perhitungan manual dengan bantuan program aplikasi Microsoft Excel menggunakan rumus KR₂₀ dari Kuder-Richardson menghasilkan angka koefisien reliabilitas keseluruhan butir soal sebesar 0,621. Berikut perhitungan reliabilitas dengan rumus KR₂₀.

S² merupakan pangkat dua dari standar deviasi. Sebelum mengitung KR₂₀ kita harus mencari S² terlebih dahulu menggunakan rumus yang dikutip dari Zainal Arifin (2013: 265) sebagai berikut.

$$S^2 = \left(\frac{n \Sigma X^2 - (\Sigma X)^2}{n (n - 1)} \right)$$

$$S^2 = \left(\frac{94 (11496) - (998)^2}{94 (94 - 1)} \right)$$

$$S^2 = 9,68$$

Setelah S² diketahui, selanjutnya menghitung jumlah proporsi siswa yang menjawab benar dikalikan dengan proporsi siswa yang menjawab salah, kemudian dijumlahkan dari 30 butir soal. Hasilnya adalah $\Sigma pq = 3,73$. Barulah KR₂₀ dapat diketahui.

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \Sigma pq}{S^2} \right)$$

$$r_{11} = \left(\frac{94}{94-1} \right) \left(\frac{9,68 - 3,73}{9,68} \right)$$

$$r_{11} = 0,621$$

Berdasarkan hasil perhitungan tersebut diketahui bahwa koefisien korelasi dari reliabilitas butir soal sebesar 0,621. Jika dilihat dari kriteria yang dikutip oleh Anas Sudijono (2015: 209) mengenai kriteria tingkat reliabilitas maka tingkat reliabilitas soal tes dapat dikategorikan kurang reliabel. Soal tes dikategorikan memiliki reliabilitas tinggi atau reliabel apabila nilai $r_{11} \geq 0,70$.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan oleh peneliti baik dalam analisis instrumen maupun analisis empiris dapat disimpulkan bahwa:

Pertama, instrumen soal berdasarkan penelaahan soal aspek materi diperoleh 26 butir soal (86,67%) sudah memenuhi seluruh aspek materi. Berdasarkan aspek konstruksi diperoleh 17 butir soal (56,67%) sudah memenuhi seluruh aspek konstruksi, dan berdasarkan aspek bahasa diperoleh sebanyak 26 butir soal (86,67%) sudah memenuhi seluruh aspek bahasa.

Kedua, instrumen soal berdasarkan aspek distribusi tingkat berpikir ranah kognitif taksonomi Bloom (revisi) diperoleh hasil distribusi butir soal, yaitu tingkat C1 33,33%, C2 20%, C3 46,67%, C4 0%, C5 0%, dan C6 0%. Hal ini menunjukkan bahwa dari 30 butir soal masih menggunakan tingkat berpikir rendah (low order thinking skill) untuk menyelesaikan soal-soal.

Ketiga, tingkat kesukaran diperoleh sebanyak 12 butir soal (40%) termasuk kategori mudah, 10 butir soal (33,33%) termasuk kategori sedang, dan 8 butir soal (26,67%) termasuk kategori sukar. Artinya, berdasarkan aspek tingkat kesukaran butir soal tes masih jauh dari kata ideal.

Keempat, daya pembeda diperoleh sebanyak 8 butir soal (26,67%) termasuk kategori baik, 11 butir soal (36,67%) termasuk kategori cukup, 10 butir soal (33,33%) termasuk kategori kurang baik, dan 1 butir soal (3,33%) termasuk kategori tidak baik. Artinya, butir soal berdasarkan aspek daya pembeda yang masuk dalam kategori baik dan cukup sudah layak untuk digunakan, sedangkan untuk butir soal kategori kurang baik dan tidak baik perlu dilakukan perbaikan.

Kelima, efektivitas pengecoh diperoleh sebanyak 20 butir soal (66,67%) memiliki jawaban pengecoh berfungsi dengan baik, dan 10 butir soal (33,33%) tersisa memiliki jawaban pengecoh yang tidak berfungsi dengan baik. Artinya, berdasarkan aspek efektivitas pengecoh pada soal tes ini sebagian besar butir soal sudah memiliki jawaban pengecoh yang berfungsi dengan baik.

Keenam, validitas butir diperoleh sebanyak 20 butir soal (66,67%) dinyatakan valid, dan 10 butir soal (33,33%) dinyatakan tidak valid. Artinya, 20 butir soal sudah layak digunakan dan 10 butir soal yang tidak valid perlu diperbaiki atau diganti dengan butir soal yang valid.

Ketujuh, reliabilitas soal diperoleh sebesar 0,621 menunjukkan bahwa tingkat reliabilitas soal kurang reliabel. Artinya, dari aspek reliabilitas, seluruh butir soal tes ini kurang dapat dipertahankan ketetapan hasil tesnya.

Saran

Dari hasil penelitian ini, peneliti memberikan, bahwa sebagai pengajar, guru seharusnya dapat membuat alat evaluasi hasil belajar dengan baik. Guru perlu memperhatikan prosedur penyusunan soal tes dari berbagai aspek dengan baik, terutama pada pilihan ganda supaya soal tes pilihan ganda yang dibuat guru memiliki kualitas yang tinggi. Sekolah sebaiknya mendukung dan mendorong para pengajar supaya memperhatikan prosedur penyusunan soal tes dari berbagai aspek dengan baik, menetapkan standar persentase pendistribusian tingkat berpikir kognitif sebagai acuan penyusunan soal tes bagi guru pengajar dan selalu mengawasi peserta tes pada pelaksanaan tes supaya tetap sesuai dengan prosedur.

DAFTAR PUSTAKA

- Arifin, Zainal. (2013). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.
- Arikunto, S. (2013). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- _____. (2016). *Dasar-Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Aksara.
- Kemenristekdikti. (2017). *Peraturan Pemerintah Nomor 19, Tahun 2017, tentang Perubahan atas Peraturan Pemerintah Nomor 74 Tahun 2008 tentang Guru*.
- Kuder, G.F. & Richardson, M.W. (1937). The Theory of The Estimation of Test Reliability. *Jurnal Psychometrika*. 2(3). Hlm. 151-160.
- Kusaeri & Suprananto. (2012). *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu.
- Maenani, L. & Oktavia, R. (2015). Analisis Butir Soal Fisika Ulangan Umum Kenaikan Kelas X Madrasah Aliyah Se-Kabupaten Banjarnegara, Jawa Tengah Tahun Pelajaran 2011/2012. *Jurnal Berkala Fisika Indonesia Universitas Ahmad Dahlan Yogyakarta*. 7(1). Hlm. 5-11.
- Siregar, T.P., Surya, E. & Syahputra, E. (2017). Quality Analysis of Multiple Choice Test and Clasical Test at X Grade Students of Senior High School. *Jurnal IJARIE University of Medan*. 3(2). Hlm. 2153-2159.
- Sugiyono. (2015). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Susetyo, B. (2015). *Prosedur Penyusunan & Analisis Tes untuk Penilaian Hasil Belajar Bidang Kognitif*. Bandung: Refika Aditama.