

ANALISIS KUALITAS BUTIR SOAL BUATAN GURU BIOLOGI KELAS X SMA NEGERI 1 TANAH PINOH

ARTIKEL PENELITIAN

Oleh

**MITA RAHMANI
NIM F05107010**



**PROGRAM STUDI PENDIDIKAN BIOLOGI JURUSAN PMIPA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
UNIVERSITAS TANJUNGPURA
PONTIANAK
2015**

**ANALISIS KUALITAS BUTIR SOAL BUATAN GURU
BIOLOGI KELAS X SMA NEGERI 1 TANAH PINOH**

ARTIKEL PENELITIAN

Oleh:

MITA RAHMANI
F05107010

Disetujui,

Pembimbing I



Dra. Hj. Kurnia Ningsih, M.Pd
NIP. 196703191991012001

Pembimbing II



Asriah Nurdini M, S.Si. M.Pd
NIP. 198105112005012002

Mengetahui,



Ketua Jurusan PMIPA



Dr. Ahmad Yani T, M.Pd
NIP. 196604011991021001

ANALISIS KUALITAS BUTIR SOAL BUATAN GURU BIOLOGI KELAS X SMA NEGERI 1 TANAH PINOH

Mita Rahmani, Kurnia Ningsih, Asriah Nurdini

Program Studi Pendidikan Biologi Jurusan Pendidikan MIPA FKIP UNTAN

Email : mita.rahmani@yahoo.com

Abstrak: Penelitian ini bertujuan untuk menganalisis validitas, reliabilitas, tingkat kesukaran, daya beda, dan distraktor butir soal buatan guru mata pelajaran biologi kelas X SMA Negeri 1 Tanah Pinoh Tahun pelajaran 2012/2013. Sampel diambil dengan teknik sampling jenuh. Instrumen dalam penelitian ini adalah lembar soal ulangan harian buatan guru mata pelajaran biologi kelas X pada SMA Negeri 1 Tanah Pinoh tahun pelajaran 2012/2013 berjumlah 20 soal pilihan ganda. Analisis validitas konstruk soal pilihan ganda dilihat dari segi materi, konstruksi dan bahasa telah sesuai, ada beberapa soal yang masih perlu diperbaiki dari segi konstruksi. Analisis validitas butir soal UH 1 menunjukkan 40% soal valid; reliabilitas sebesar 0,76 artinya memiliki keajegan tinggi; tingkat kesukaran butir soal 40% sukar, 55% sedang dan 5% soal mudah; daya beda 45% jelek, 35% cukup dan 20% baik; 80% pengecoh berfungsi. Pada UH 2 60% soal valid; reliabilitas sebesar 0,78 artinya memiliki keajegan tinggi; 35% soal sukar dan 65% soal sedang; daya beda 30% jelek, 45% cukup dan 25% baik; 75% pengecoh berfungsi.

Kata kunci: Kualitas Soal, Butir Soal, Pelajaran Biologi

Abstract: This research aims to analyze the validity, reliability, level of difficulty, discrimination power, of the test items. The test items were made by the Biology teacher in Tenth grade of SMA N 1 Tanah Pinoh year 2012/2013. The samples were taken by using the stratified sampling. The instrument of this research was the test items of the daily examination of the tenth grade of SMA N 1 Tanah Pinoh which consist of 20 multiple choice items. The validity analysis of the multiple choices could be previewed by the material aspect, form and the language features used in the test items. Even, there were some needed to prove in the test items language used. The researcher found that the validity analysis of the test items in UH 1 showed that about 40% were valid, the reliability was 0,76, the level of difficulty was 40% hard, 55% medium, and 5% easy, the discrimination power 45% was bad, 35% was enough, and 20% was good and the obstacle in test items was about 80%. While in UH 2, the validity was 60% was valid, reliability was 0,78, the level of difficulty was 35% hard, 65 medium the discrimination power was about 30% bad, 45 % enough, and 25% good while the obstacle was about 75%.

Key words: *The Test Item Quality, Test Items, Biology*

Proses pembelajaran merupakan sistem yang terdiri atas beberapa komponen yang saling berkaitan dan berinteraksi dalam mencapai tujuan pembelajaran. Salah satu komponen yang terpenting dalam proses belajar mengajar adalah evaluasi. Evaluasi dalam bahasa Inggris dikenal dengan istilah *evaluation* adalah suatu proses sistematis untuk menentukan atau membuat keputusan sampai sejauh mana tujuan program telah dicapai (Gronlund dalam Djaali dan Muljono, 2008: 1). Dalam kaitannya dengan pendidikan, Nurkancana dan Sunartana (1986: 1) mengemukakan bahwa evaluasi pendidikan merupakan suatu tindakan atau suatu proses untuk menentukan nilai terhadap segala sesuatu yang berkaitan dalam dunia pendidikan.

Dalam kegiatan evaluasi diperlukan alat atau teknik penilaian, sehingga pelaksanaannya akan lebih terarah. Alat evaluasi dalam pendidikan yang digunakan untuk mengumpulkan data dapat berupa tes atau nontes (Purwanto, 2011: 56). Nurkancana dan Sunartana (1986: 25) menyatakan tes adalah suatu cara untuk mengadakan penilaian yang berbentuk suatu tugas atau serangkaian tugas yang harus dikerjakan oleh siswa atau sekelompok siswa sehingga menghasilkan nilai tentang tingkah laku atau prestasi siswa sebagai peserta didik. Tes dapat disusun berupa tes (soal) berbentuk objektif atau subjektif. Menurut Purwanto (2011: 72) tes objektif adalah tes yang keseluruhan informasi yang diperlukan untuk menjawab tes telah tersedia. Tes subjektif menurut Nurkancana dan Sunartana (1986: 42) merupakan suatu bentuk tes yang terdiri dari pertanyaan atau suruhan yang menghendaki jawaban berupa uraian-uraian yang relatif panjang. Dua bentuk tes yang digunakan dalam evaluasi ini harus dapat dipertanggungjawabkan, artinya bahwa tes tersebut dapat memenuhi syarat sebagai alat evaluasi yang baik bila dilihat dari kualitas butir soal.

Untuk mengetahui kualitas butir soal dapat diungkap melalui analisis butir soal. Menurut Nurgiantoro (dalam Sastrawati, 2005: 2) analisis butir soal dapat memberikan informasi secara terinci tentang keadaan masing-masing butir soal seperti kekuatan dan kelemahan butir soal, spesifikasi soal secara lengkap dan masalah yang terkandung dalam soal seperti kesalahan dalam pembuatan kunci jawaban, soal terlalu sukar atau mudah dan lain sebagainya. Soal yang telah disusun oleh guru baik soal ulangan akhir semester ataupun soal ulangan harian perlu dilakukan analisis. Hal tersebut mengacu pada pendapat Suryabrata (1987: 3) seperti yang dikutip berikut: Soal-soal yang telah ditulis dengan hati-hati berdasarkan berbagai pertimbangan tidak dapat begitu saja dianggap soal-soal yang sudah baik. Soal-soal itu masih perlu diuji melalui penelaahan soal.

Baik buruknya suatu tes atau alat evaluasi dapat ditinjau dari validitas, reliabilitas, tingkat kesukaran dan daya beda (Nurkancana dan Sunartana, 1986: 127). Sebuah tes disebut valid atau memiliki validitas apabila tes itu dapat tepat mengukur apa yang hendak diukur. Validitas butir perlu dicari untuk mengetahui butir-butir tes manakah yang menyebabkan soal secara keseluruhan jelek karena memiliki validitas rendah. Lebih lanjut Butir soal dikatakan valid apabila memiliki dukungan besar terhadap skor total. Skor pada butir soal menyebabkan skor total menjadi tinggi atau rendah. Dengan kata lain dapat dikatakan bahwa butir soal memiliki validitas yang tinggi jika skor pada butir soal memiliki kesejajaran dengan skor total. Kesejajaran ini diartikan dengan korelasi, sehingga

untuk mengetahui validitas butir soal digunakan rumus korelasi (Arikunto, 2008: 76). Dalam penelitian ini, dilakukan analisis validitas butir soal dengan menggunakan rumus korelasi *product moment* angka kasar.

Reliabilitas adalah ketetapan suatu tes apabila diteskan pada subjek yang sama (Arikunto, 2008: 90). Untuk mengetahui ketetapan ini pada dasarnya dilihat kesejajaran hasil. Untuk menganalisis reliabilitas soal tes dalam penelitian ini digunakan rumus *Kuder-Richarson 20* atau disingkat KR-20.

Bermutu atau tidaknya butir-butir soal tes hasil belajar juga dapat diketahui dari derajat kesukaran atau tingkat kesulitan yang dimiliki masing-masing butir soal tersebut. Butir-butir soal tes hasil belajar dapat dikatakan sebagai butir soal yang baik apabila butir soal tersebut tidak terlalu sukar dan tidak pula terlalu mudah dengan kata lain, derajat kesukaran butir soal adalah sedang atau cukup (Sudijono, 2007: 370). Menurut Arikunto (2008: 207) Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sulit. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena diluar jangkauannya. Bilangan yang menunjukkan sukar dan mudahnya suatu soal disebut indeks kesukaran (*difficulty index*). Besarnya indeks kesukaran antara 0,0 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal. Soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal tersebut terlalu sukar, sebaliknya indeks 1,0 menunjukkan soal tersebut terlalu mudah. Di dalam istilah evaluasi, indeks kesukaran ini diberi simbol P (singkatan dari Proporsi).

Butir soal tes yang baik juga harus dapat menunjukkan daya pembedanya. Menurut Arikunto (2008: 211), “daya beda adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang (berkemampuan rendah)”. Menurut Anastasi dan Urbina (dalam Purwanto, 2011: 102), daya beda berhubungan dengan derajat kemampuan butir membedakan dengan baik perilaku pengambil tes dalam tes yang dikembangkan. Soal dapat dikatakan mempunyai daya pembeda jika soal tersebut dapat dijawab oleh siswa berkemampuan tinggi dan tidak dapat dijawab oleh siswa berkemampuan rendah. Jika suatu soal dapat dijawab oleh siswa pintar maupun kurang, berarti soal tersebut tidak mempunyai daya beda, demikian juga jika soal tersebut tidak dapat dijawab oleh siswa pintar dan siswa kurang, berarti soal tersebut tidak baik sebab tidak mempunyai daya pembeda. Menurut Sudijono (2007: 387), daya pembeda butir soal dapat diketahui dengan melihat besar kecilnya angka indeks deskriminasi butir soal. Angka indeks deskriminasi butir soal adalah sebuah angka atau bilangan yang menunjukkan besar kecilnya daya pembeda (*discriminatory power*) yang dimiliki oleh sebutir soal. *Discriminatory power* pada dasarnya dihitung atas dasar pembagian siswa kedalam dua kelompok, yaitu: kelompok atas (*the higher group*) – yakni kelompok siswa yang tergolong pandai, dan kelompok bawah (*the lower group*) – yakni kelompok siswa yang tergolong kurang pandai. Indeks diskriminasi butir soal umumnya diberi lambang dengan huruf D.

Setiap butir soal pada tes objektif bentuk pilihan ganda terdiri dari beberapa kemungkinan jawaban, atau yang lebih dikenal dengan option atau alternatif.

Option biasanya berjumlah tiga atau lima buah, dan dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir soal itu salah satunya adalah jawaban betul (kunci jawaban) sedangkan sisanya merupakan jawaban salah. Jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (distraktor = pengecoh). Tujuan pemasangan distraktor pada setiap butir item adalah agar dari sekian banyak siswa mengikuti tes ada yang tertarik memilihnya, sebab mereka menyangka bahwa distraktor yang mereka pilih merupakan jawaban betul. Semakin banyak siswa terkecoh, maka distraktor makin dapat menjalankan fungsinya sebaik-baiknya. Sebaliknya, jika distraktor yang dipasang tidak ada yang memilih, maka distraktor tidak dapat menjalankan fungsinya dengan baik. Menurut Sudijono (2007: 411) “distraktor dinyatakan telah berfungsi dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih 5% dari seluruh peserta tes”.

Penelitian mengenai analisis soal yang dibuat oleh guru pernah dilakukan oleh Ariyana (2011) menunjukkan rata-rata persentase tingkat kesukaran mudah 8%, sedang 70%, dan sukar 22%; reliabilitas sebesar 0,711; daya pembeda soal 26% baik; efektifitas pengecoh sebanyak 18% kurang berfungsi. Penelitian yang dilakukan oleh Afyana (2010) mengenai analisis soal latihan UN IPA di Kabupaten Batang menunjukkan kualitas soal dalam kategori kurang baik karena soal didominasi oleh tingkat kesukaran soal yang tinggi.

Analisis terhadap soal ulangan harian buatan guru mata pelajaran biologi kelas X SMA Negeri 1 Tanah Pinoh belum dilakukan karena tidak ada tuntutan dari pihak sekolah. Dari penelitian ini diharapkan bisa diketahui kualitas soal buatan guru di sekolah tersebut sehingga dapat menjadi umpan balik untuk perbaikan sistem evaluasi dikelas pada periode berikutnya serta menjadi masukan untuk perbaikan program evaluasi di sekolah. Oleh karena itu, perlu dilakukan analisis terhadap soal buatan guru mata pelajaran biologi kelas X SMA Negeri 1 Tanah Pinoh tahun pelajaran 2012/2013 supaya diketahui kualitas soal yang terdiri dari validitas konstruk, validitas butir soal, reliabilitas, tingkat kesukaran, daya beda dan distraktor soal ulangan harian buatan guru tersebut.

METODE

Bentuk penelitian yang digunakan dalam penelitian ini adalah deskriptif. Sampel dalam penelitian ini adalah soal ulangan harian buatan guru mata pelajaran Biologi kelas X SMA Negeri 1 Tanah Pinoh tahun pelajaran 2012/2013 dengan bentuk soal pilihan ganda terdiri dari soal ulangan harian materi 1 berjumlah 20 butir dan soal ulangan harian materi 2 berjumlah 20 butir. Prosedur dalam penelitian ini terdiri dari 2 tahap, yaitu: 1) tahap persiapan dan 2) tahap pelaksanaan.

Tahap persiapan

Langkah-langkah dalam tahap persiapan, antara lain: (1) mengadakan observasi awal untuk memperoleh informasi mengenai pelaksanaan ulangan harian; (2) meminta izin penelitian ke SMA Negeri 1 Tanah Pinoh sebagai tempat pelaksanaan penelitian; (3) membuat format penelaahan soal bentuk pilihan ganda untuk persiapan analisis validitas konstruk soal.

Tahap Pelaksanaan

Langkah-langkah dalam tahap pelaksanaan, antara lain: (1) Mengumpulkan data penelitian yaitu: lembar soal ulangan harian buatan guru mata pelajaran Biologi kelas X tahun pelajaran 2012/2013 (soal ulangan harian materi 1) dan kunci jawaban serta mengumpulkan lembar jawaban siswa kelas X peserta ulangan harian; (2) Menganalisis soal ulangan harian buatan guru (soal ulangan harian materi 1) dengan cara menghitung validitas, reliabilitas, tingkat kesukaran, daya pembeda dan distraktor butir soal; (3) Membuat interpretasi data hasil perhitungan validitas, reliabilitas, tingkat kesukaran, daya pembeda dan distraktor butir soal; (4) Berdasarkan pada hasil analisis dan penelaahan, dilakukan diskusi dengan guru untuk pembuatan soal ulangan harian pada materi selanjutnya; (5) Mengumpulkan lembar soal ulangan harian buatan guru mata pelajaran Biologi kelas X tahun pelajaran 2012/2013 (soal ulangan harian materi 2) dan kunci jawaban serta mengumpulkan lembar jawaban siswa peserta ulangan harian; (6) Menganalisis soal ulangan harian buatan guru (soal ulangan harian materi 2) dengan cara menghitung validitas butir soal, reliabilitas soal, tingkat kesukaran butir soal, daya pembeda butir soal dan distraktor soal ulangan harian buatan guru serta menelaah soal pilihan ganda; (7) Membuat interpretasi data hasil perhitungan validitas, reliabilitas, tingkat kesukaran, daya pembeda dan distraktor butir soal; (8) Penelaahan terhadap soal ulangan harian buatan guru mata pelajaran biologi kelas X SMA Negeri 1 Tanah Pinoh (soal ulangan harian materi 1 dan materi 2) dilakukan dengan bantuan 3 orang guru biologi. Penelaahan dilakukan dengan menggunakan format penelaahan soal bentuk pilihan ganda.

Tahap Akhir

Langkah-langkah dalam tahap akhir, antara lain: (1) Membuat kesimpulan sebagai jawaban dari rumusan masalah; (2) Menyusun laporan penelitian.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Dari hasil analisis yang telah dilakukan diketahui validitas konstruk, validitas butir soal, reliabilitas, tingkat kesukaran, daya pembeda, dan distraktor. Data yang dianalisis dalam penelitian ini adalah soal ulangan harian 1 (UH 1) yaitu materi klasifikasi tumbuhan yang selanjutnya disebut UH 1 dan soal ulangan harian 2 (UH 2) yaitu materi klasifikasi hewan yang selanjutnya disebut UH 2 serta lembar jawaban siswa. Soal dibuat oleh guru mata pelajaran Biologi kelas X SMA Negeri 1 Tanah Pinoh Tahun Pelajaran 2012/2013. Soal ulangan harian disajikan dalam bentuk tes pilihan ganda berjumlah 20 butir soal dengan lima pilihan jawaban (A, B, C, D, dan E). Hasil analisis yang diperoleh dapat dilihat sebagai berikut:

Validitas Konstruk

Validitas konstruk soal dianalisis dengan melakukan penelaahan terhadap soal pilihan ganda yang telah diberikan pada UH 1 dan UH 2 dengan melibatkan 3 orang guru mata pelajaran biologi yang ada di Kota Pontianak. Berdasarkan hasil

analisis validitas konstruk soal UH 1 diketahui ada tiga soal yakni soal nomor 1, 15 dan 20 tidak sesuai pada aspek penelaahan aspek 5 yaitu pokok soal dirumuskan dengan singkat, jelas dan tegas. Sedangkan pada hasil analisis validitas konstruk soal UH 2 diketahui satu soal yaitu soal nomor 6 tidak sesuai pada aspek 13 yaitu pilihan jawaban berbentuk angka/waktu disusun berdasarkan urutan besar kecilnya angka atau kronologisnya. Sehingga dapat dikatakan soal cukup baik bila dilihat dari segi validitas konstruk soal karena sebagian besar soal telah sesuai dengan aspek penelaahan meskipun ada beberapa soal yang masih perlu diperbaiki pada aspek konstruksi soal.

Validitas Butir Soal

Berdasarkan hasil analisis validitas butir soal UH 1 dan UH 2 dapat diketahui validitas butir soal pada UH 1 kategori sangat rendah sebanyak 20%, kategori rendah sebanyak 40%, kategori cukup 35% dan kategori tinggi 5% dari 20 soal. Sedangkan pada soal UH 2 kategori sangat rendah 15%, kategori rendah 25%, kategori cukup 55% dan kategori tinggi 5% dari 20 soal. Tabel berikut menunjukkan persentase soal yang valid (yaitu soal yang memiliki validitas kategori cukup, tinggi dan sangat tinggi) dan persentase soal yang tidak valid (soal yang memiliki validitas kategori rendah dan sangat rendah).

Tabel 1. Persentase Soal Valid dan Soal yang Tidak Valid pada UH 1 dan UH 2

UH	Kategori	Jumlah	Persentase	No. Soal
UH 1	Tidak valid	12	60%	1, 4, 7, 8, 9, 10, 11, 12, 16, 17, 19, 20
	Valid	8	40%	2, 3, 5, 6, 13, 14, 15, 18
UH 2	Tidak valid	8	40%	3, 4, 6, 9, 10, 12, 15, 16
	Valid	12	60%	1, 2, 5, 7, 8, 11, 13, 14, 17, 18, 19, 20

Reliabilitas

Berdasarkan analisis reliabilitas soal pilihan ganda buatan guru SMA Negeri 1 Tanah Pinoh pada UH 1 dan UH 2 dapat dirangkum sebagai berikut:

Tabel 2. Rangkuman Hasil Perhitungan Reliabilitas Soal pada UH 1 dan UH 2

UH	Reliabilitas	Interpretasi
1	0,76	Tinggi
2	0,78	Tinggi

Dari tabel di atas diketahui bahwa reliabilitas soal UH 1 sebesar 0,76 dengan kategori tinggi dan reliabilitas soal UH 2 sebesar 0,78 kategori tinggi sehingga bisa dikatakan soal tersebut dilihat dari segi reliabilitas baik.

Tingkat kesukaran

Berdasarkan analisis tingkat kesukaran butir soal diketahui bahwa pada UH 1 sebanyak 40% soal kategori sukar, 55% soal kategori sedang dan 5% soal kategori mudah. Sedangkan pada UH 2 diketahui sebanyak 35% soal kategori sukar dan 65% soal kategori sedang. Hal tersebut dapat dirangkum sebagai berikut.

Tabel 3. Rangkuman Kategori Tingkat Kesukaran pada UH 1 dan UH 2

Kategori tingkat kesukaran	UH 1			UH 2		
	Jumlah	Persentase (%)	No. Soal	Jumlah	Persentase (%)	No. Soal
Sukar	8	40	2, 4, 7, 10, 15, 17, 19, 20	7	35	3, 7, 10, 12, 15, 16, 20
Sedang	11	55	3, 5, 6, 8, 9, 11, 12, 13, 14, 16, 18	13	65	1, 2, 4, 5, 6, 8, 9, 11, 13, 14, 17, 18, 19
Mudah	1	5	1	-	-	-

Daya pembeda

Berdasarkan analisis daya pembeda butir soal UH 1 dan UH 2 diketahui bahwa soal UH 1 dengan daya beda kategori baik berjumlah 20%, kategori cukup sebanyak 40% dan kategori jelek sebanyak 40%. Pada UH 2, sebanyak 25% soal kategori baik, 45% soal kategori cukup dan 30% soal memiliki daya beda butir soal kategori jelek. Berikut rangkuman kategori daya beda :

Tabel 4. Rangkuman Kategori Daya Pembeda pada UH 1 dan UH 2

Kategori daya pembeda	UH 1			UH 2		
	No. Soal	Jumlah soal	Persentase (%)	No. Soal	Jumlah soal	Persentase (%)
Jelek	1, 4, 7, 9, 11, 17, 19, 20	8	40	3, 10, 12, 15, 16, 20	6	30
Cukup	2, 5, 8, 10, 12, 15, 16, 18	8	40	4, 5, 6, 7, 9, 11, 14, 18, 19	9	45
Baik	3, 6, 13, 14	4	20	1, 2, 8, 13, 17	5	25

Distraktor

Hasil analisis distraktor soal pilihan ganda UH 1 dan UH 2 diketahui bahwa dari 20 soal yang disajikan 80% pengecoh berfungsi dan 20% pengecoh tidak berfungsi. Sedangkan pada UH 2, 75% pengecoh berfungsi dan 25% pengecoh tidak berfungsi. Sehingga bisa dikatakan soal cukup baik bila dilihat dari distraktor soal karena sebagian besar pengecoh telah berfungsi. Hal ini dapat dirangkum sebagai berikut.

Tabel 5. Rangkuman Kategori Pengecoh pada UH 1 dan UH 2

Kategori	UH 1			UH 2		
	No. Soal	Jumlah soal	Persentase (%)	No. Soal	Jumlah soal	Persentase (%)
Berfungsi	2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20	16	80	2, 3, 5, 6, 8, 9, 10, 11, 12, 14, 16, 17, 18, 19, 20	15	75
Tidak berfungsi	1, 8, 12, 17	4	20	1, 4, 7, 13, 15	5	25

Pembahasan

Validitas Konstruk Soal

Analisis validitas konstruk soal pilihan ganda dilakukan dengan menggunakan format penelaahan soal pilihan ganda yang ditelaah oleh 3 orang guru biologi SMA di Kota Pontianak. Penelaahan dilakukan dengan menggunakan format penelaahan soal bentuk pilihan ganda yang dikutip dari Wahidmuri dkk (2010: 124). Analisis validitas konstruk berkaitan dengan susunan atau tampilan soal. Adapun 3 aspek yang ditelaah pada analisis validitas konstruk soal pilihan ganda meliputi aspek materi, konstruksi dan bahasa.

Aspek yang ditelaah pada aspek materi meliputi (1) kesesuaian soal dengan kompetensi dasar (menuntut tes tertulis untuk bentuk pilihan ganda); (2) materi yang ditanyakan sesuai dengan kompetensi (urgensi, relevansi, kontinuitas, keterpakaian sehari-hari tinggi); (3) pilihan jawaban homogen dan logis; (4) hanya ada satu kunci jawaban.

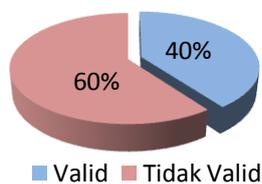
Pada aspek konstruksi, aspek yang ditelaah meliputi: (1) pokok soal harus dirumuskan dengan singkat, jelas dan tegas; (2) rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja; (3) pokok soal tidak memberi petunjuk kunci jawaban; (4) pokok soal bebas dari pernyataan yang bersifat negatif ganda; (5) pilihan jawaban homogeny dan logis dilihat dari segi materi; (6) gambar, grafik, tabel, diagram atau sejenisnya jelas; (7) panjang pilihan jawaban harus relatif sama; (8) pilihan jawaban tidak menggunakan pernyataan "semua jawaban diatas salah/benar" atau sejenisnya; (9) pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya angka atau kronologisnya; dan (10) butir soal tidak tergantung pada jawaban soal yang sebelumnya

Aspek bahasa/budaya terdiri dari: (1) menggunakan bahasa yang sesuai dengan kaidah Bahasa Indonesia; (2) menggunakan bahasa yang komunikatif; (3) tidak menggunakan bahasa yang berlaku setempat dan (4) pilihan jawaban tidak mengulang kata atau kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian.

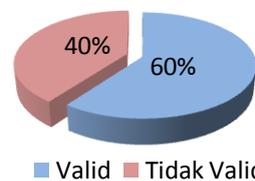
Berdasarkan hasil analisis, diketahui bahwa dilihat dari segi materi, soal tersebut 100% soal pada UH 1 telah sesuai dengan aspek penelaahan. Jika dilihat dari segi konstruksi, masih ada beberapa soal yang perlu direvisi karena tidak sesuai dengan beberapa aspek penelaahan. Aspek tersebut antara lain aspek no 5, yaitu pokok soal yang disusun tidak dirumuskan dengan singkat, jelas dan tegas menurut salah satu atau dua dari guru yang menelaah, soal yang dimaksud adalah soal nomor 1, 15 dan 20. Dilihat dari segi bahasa, soal pilihan ganda UH 1 secara keseluruhan sudah sesuai dengan aspek penelaahan. Pada UH 2 pada aspek materi semua soal telah sesuai. Pada aspek konstruksi, soal nomor 6 tidak sesuai dengan aspek ke 13 yaitu pilihan jawaban yang berbentuk angka/waktu yang disusun berdasarkan urutan besar kecilnya angka atau kronologisnya menurut ketiga guru yang menelaah. Dilihat dari segi bahasa, soal pilihan ganda UH 2 secara umum sudah sesuai dengan aspek penelaahan. Dari hasil analisis tersebut, soal UH 1 dan UH 2 bisa dikatakan telah memiliki validitas konstruk meskipun ada beberapa soal yang masih perlu di perbaiki pada aspek konstruksi, namun secara umum soal telah cukup baik.

Validitas Butir Soal

Berdasarkan hasil analisis validitas butir soal pada UH 1 diketahui bahwa korelasi antara skor item terhadap skor total ulangan harian secara keseluruhan rendah. Dari 20 soal yang disajikan, sebanyak 60% atau 12 butir soal tidak valid sedangkan soal yang valid hanya sebanyak 40% atau 8 butir. Soal-soal yang tidak valid adalah soal nomor: 1, 4, 7, 8, 9, 10, 11, 12, 16, 17, 19, 20. Pada UH 2 diketahui bahwa validitas soal secara keseluruhan cukup tinggi. Dari 20 soal yang disajikan, sebanyak 60% atau 12 butir soal valid sedangkan soal yang tidak valid sebanyak 40% atau 8 butir. Soal-soal yang tidak valid adalah soal nomor: 3, 4, 6, 9, 10, 12, 15, 16.



Gambar 1: Hasil Analisis Validitas Butir Soal UH 1



Gambar 2: Hasil Analisis Validitas Butir Soal UH 2

Soal dikatakan valid atau memiliki validitas tinggi merupakan soal yang dapat mengukur kompetensi yang diharapkan. Sedangkan soal yang tidak valid atau memiliki validitas rendah artinya soal tersebut tidak bisa mengukur kompetensi yang diharapkan. Menurut Arikunto (2008: 65) suatu teknik evaluasi dikatakan mempunyai validitas yang tinggi (disebut valid) jika teknik evaluasi

tersebut dapat sepenuhnya mengukur kemampuan tertentu yang diharapkan. Sehingga dapat dikatakan bahwa alat tes tersebut mampu untuk mengukur kemampuan siswa dalam pencapaian kompetensi yang diharapkan pada materi tersebut. Skor pada item/butir soal menyebabkan skor total menjadi tinggi atau rendah dengan kata lain sebuah item memiliki validitas yang tinggi jika skor pada item mempunyai kesejajaran dengan skor total.

Pada analisis butir, butir soal akan dilihat karakteristiknya dan dipilih butir-butir yang baik. Butir yang baik adalah butir-butir yang karakteristiknya memenuhi syarat sebagaimana kriteria butir tes yang baik. Analisis butir dilakukan atas sejumlah banyak butir tes. Analisis butir akan menggugurkan sebagian butir yang dianalisis karena tidak memenuhi kemampuan mengukur hasil belajar dengan baik (Purwanto, 2011: 97). Soal yang mempunyai validitas rendah atau tidak valid artinya soal tersebut tidak dapat mengukur kompetensi yang diharapkan. Pada soal UH 1, soal yang memiliki validitas butir yang kategori sangat rendah terdiri dari soal nomor 4, nomor 7, nomor 11 dan nomor 20, dan ada 8 soal memiliki validitas butir soal kategori rendah, yaitu soal nomor 1, 8, 9, 10, 12, 16, 17, dan 19. Pada soal UH 2, soal yang memiliki validitas butir soal kategori sangat rendah ada tiga soal, yaitu soal nomor 3, 10 dan 15, dan kategori rendah ada 5 soal yaitu soal nomor 4, 6, 9, 12 dan 16.

Reliabilitas

Dari hasil perhitungan dengan rumus KR-20 diketahui nilai indeks reliabilitas. Indeks reliabilitas berkisar antara 0 – 1 dengan 5 kriteria. Semakin tinggi koefisien reliabilitas suatu tes, semakin tinggi pula keajegan atau ketepatannya. Nilai indeks reliabilitas pada UH 1 sebesar 0,76 dan pada UH 2 sebesar 0,78 dengan arti soal tersebut memiliki reliabilitas yang tinggi. Reliabilitas tinggi yang dimaksud dalam hal ini meliputi ketepatan atau kecermatan hasil pengukuran dan keajegan atau kestabilan dari hasil pengukuran.

Tingkat Kesukaran

Tingkat kesukaran butir soal tes menunjukkan seberapa sulit atau mudahnya butir-butir soal tes dan tes secara keseluruhan yang telah diselenggarakan. Indeks tingkat kesukaran merupakan rasio antara penjawab item dengan benar dan banyaknya penjawab item (Gronlund dalam Santos, 2012: 5). Analisis tingkat kesukaran menunjukkan apakah butir soal tes itu tergolong terlalu sulit, sulit, sedang, mudah atau terlalu mudah. Melalui kajian dan analisis tingkat kesukaran terhadap tes yang diujikan akan dapat diungkap kelayakan soal tes, baik masing-masing butir soal tes maupun keseluruhan soal tes.

Menurut Sudjana (2004: 136), sebaiknya sebuah paket soal yang diberikan kepada siswa memiliki keseimbangan antara sukar : cukup : mudah dengan perbandingan 3:4:3 atau 2:5:3. Dari 20 soal berdasarkan tabel 3 diketahui bahwa pada UH 1 sebanyak 8 soal kategori sukar, 11 soal kategori cukup dan 1 soal kategori mudah dengan perbandingan sukar 8: cukup 11: mudah 1 atau 4:5,5:0,5. Sedangkan pada UH 2 diketahui sebanyak 7 soal kategori sukar dan 13 soal kategori cukup dengan perbandingan sukar 7: cukup 13: mudah 0 atau . 3,5:6,5:0. Hasil perbandingan tersebut menunjukkan soal-soal pilihan ganda soal UH 1 dan

UH 2 memiliki proporsi yang tidak seimbang. Dengan kata lain bisa dikatakan soal didominasi oleh tingkat kesukaran cukup pada UH 1 dan UH 2. Dalam penyusunan soal, perlu diperhatikan besarnya persentase tingkat kesukaran soal. Menurut Joesmani (1988) tingkat kesukaran antara 25-75% (kategori cukup) merupakan tingkat kesukaran yang memadai. Makin rendah angka persentase tingkat kesukaran soal, maka soal tersebut semakin sukar, sebab sedikit peserta tes yang menjawab benar soal tersebut. Soal yang baik menurut Arikunto (2008: 207) adalah soal yang tidak terlalu mudah atau terlalu sukar. Soal yang mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya, soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba karena diluar jangkauannya.

Daya Pembeda

Menurut Arikunto (2008:211), daya pembeda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang pandai dengan siswa yang kurang pandai. Suatu soal yang dapat dijawab benar oleh siswa yang pandai ataupun siswa yang kurang pandai, maka soal tersebut dikatakan soal yang tidak baik karena tidak mempunyai daya pembeda.

Berdasarkan hasil analisis daya beda soal-soal pilihan ganda soal UH 1 dengan daya beda kategori baik berjumlah 20%, kategori cukup sebanyak 40% dan kategori jelek sebanyak 40%. Pada UH 2, sebanyak 25% soal kategori baik, 45% soal kategori cukup dan 30% soal memiliki daya beda butir soal kategori jelek. Jika dilihat dari daya beda, soal UH 1 dan UH 2 bisa dikatakan baik karena sebagian besar soal memiliki daya beda yang cukup baik dimana soal yang memiliki daya beda baik berarti soal tersebut dapat membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah. Sehingga dapat dikatakan bahwa pada UH 1 dan UH 2 memiliki daya beda yang cukup baik karena sebagian besar memiliki daya beda yang baik, artinya soal tersebut sudah cukup dapat membedakan antara siswa yang berkemampuan tinggi dan berkemampuan rendah. Artinya adalah siswa yang pandai lebih banyak menjawab benar soal yang bersangkutan, sedangkan siswa yang kurang pandai lebih banyak menjawab salah.

Meskipun demikian, ada beberapa soal yang masih memiliki daya beda rendah pada UH 1, yaitu soal nomor 1, 4, 7, 9, 11, 17, 19 dan 20.

Pada soal nomor 7 daya beda bertanda negatif yaitu sebesar -0,09 dan pada soal nomor 11 sebesar -0,06. Menurut Sudijono (2007:388), apabila angka indeks deskriminasi butir soal bertanda negatif artinya bahwa butir soal yang bersangkutan lebih banyak dijawab benar oleh siswa kelompok bawah dibandingkan siswa kelompok atas, atau siswa yang termasuk kelompok pandai lebih banyak menjawab salah sedangkan siswa kelompok kurang pandai justru lebih banyak jawabannya benar.

Distraktor

Berdasarkan tabel 5. yang menunjukkan hasil analisis distraktor soal pilihan ganda pada UH 1 diketahui bahwa 16 soal memiliki pilihan jawaban pada soal berfungsi. Artinya pilihan jawaban (bukan kunci jawaban) telah berfungsi sebagaimana mestinya yaitu sebagai pengecoh. Sisanya, yaitu sebanyak 4 soal

yaitu soal nomor 1 pilihan A, C, D; soal nomor 8 pilihan C; soal nomor 12 pilihan E; soal nomor 17 pilihan D pengecoh tidak berfungsi. Pada UH 2 diketahui bahwa 15 soal memiliki pilihan jawaban pada soal berfungsi. Artinya pilihan jawaban telah berfungsi sebagai pengecoh. Sisanya, yaitu sebanyak 5 soal yaitu soal nomor 1 pilihan B; soal nomor 2 pilihan D; soal no 7 pilihan C, soal no 13 pilihan E dan 15 pilihan E belum berfungsi dengan baik sebagai pengecoh.

Menurut Aprianto (2008) terdapat beberapa faktor yang mempengaruhi berfungsi tidaknya pengecoh yaitu jika soal terlalu mudah, pokok soal memberi petunjuk pada kunci jawaban dan siswa sudah mengetahui materi yang akan ditanyakan terlalu mudah. Tes pilihan ganda yang disusun tanpa memperhatikan homogenitas tidaknya pilihan jawaban akan berpeluang untuk tidak berfungsi. Karena peserta tes akan dengan mudah menebak tanpa berfikir panjang akan langsung menjawab pada kunci jawaban, artinya tidak menghiraukan pilihan jawaban lain sebagai pengecoh yang tidak homogen. Demikian juga jika pokok soal memberi petunjuk untuk jawaban yang benar. Petunjuk untuk pilihan jawaban yang benar membuat peserta tes menjawab sesuai dengan petunjuk. Hal ini akan membuat alternatif jawaban lain tidak berfungsi.

Pengecoh dikatakan berfungsi jika dipilih oleh sebagian besar siswa yang berkemampuan rendah dan dipilih minimal 5% dari seluruh peserta tes yang berkemampuan tinggi. Jika pengecoh lebih banyak dipilih oleh siswa yang berkemampuan tinggi maka dapat dikatakan pengecoh tersebut menyesatkan. Apabila pengecoh dipilih secara merata oleh peserta tes, maka pengecoh tersebut berfungsi. Menurut Widodo (2010), penyebab pengecoh yang tidak dipilih oleh peserta tes karena terlalu kelihatan menyesatkan. Selain itu, perlu diperhatikan lagi apakah pilihan jawaban tidak homogen atau justru siswa sudah benar-benar memahami konsep materi yang diajarkan.

Sebagai tindak lanjut atas hasil analisis terhadap berfungsi tidaknya pengecoh, maka untuk pengecoh yang telah berfungsi pada soal tersebut dapat digunakan pada ulangan harian semester selanjutnya sedangkan pengecoh yang belum berfungsi perlu diganti atau direvisi.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan hasil penelitian dan pembahasan, disimpulkan bahwa validitas butir soal pada UH 1 40% soal valid dan pada UH 2 60% soal valid; Reliabilitas soal UH 1 sebesar 0,76 dan UH 2 sebesar 0,78 dengan kategori tinggi; Tingkat kesukaran soal UH 1 40% soal kategori sukar, 55% soal kategori sedang, 5% soal kategori mudah. Pada UH 2 35% soal kategori sukar dan 65% soal kategori sedang; Daya beda soal UH 1, 20% soal kategori baik, 40% kategori cukup dan 40% kategori jelek. Pada UH 2, 25% soal kategori baik, 45% soal kategori cukup dan 30% soal kategori jelek; Distraktor pada soal UH 1 sebanyak 80% pengecoh berfungsi dan pada soal UH 2 75% pengecoh berfungsi.

Saran

Sebaiknya guru mata pelajaran melaksanakan kegiatan analisis terhadap soal ulangan yang telah dibuat, agar diketahui kualitas soal sehingga kedepannya dibuat lebih baik lagi.

DAFTAR RUJUKAN

- Afiana. 2010. *Analisis kualitas Soal Latihan UN Mata Pelajaran IPA menggunakan Program ITEMAN di Batang*. skripsi. Semarang: Universitas Negeri Semarang.
- Aprianto. 2008. *Kualitas Butir Tes Soal Pilihan Ganda*. Banjarmasin. *On line at <http://apri76.wordpress.com/2008/12/31/kualitas-butir-tes-soal-pilihan-ganda>* (diakses tanggal 25 Maret 2013).
- Arikunto, Suharsimi. 2008. *Dasar-dasar Evaluasi Pendidikan*. (Edisi Revisi). Jakarta: Bumi Aksara.
- Ariyana, Lilis Tri. 2011. *Analisis Butir Soal Ulangan Akhir Semester Gasal IPA kelas IX SMP di Kabupaten Grobogan*. skripsi. Semarang: Universitas Negeri Semarang.
- Djaali & Muljono, Pudji. 2008. *Pengukuran Dalam Bidang Pendidikan*. Jakarta: Grasindo.
- Joesmani. 1988. *Pengukuran dan Evaluasi dalam Pengajaran*. Jakarta: Departemen Pendidikan dan Kebudayaan Direktorat Jendral Pendidikan Tinggi Proyek Pengembangan Lembaga Pendidikan Tenaga Kependidikan.
- Nurkancana, Wayan & Sunartana, P.P.N. 1986. *Evaluasi Pendidikan*. (Cetakan ke-4). Surabaya: Usaha Nasional.
- Purwanto. 2011. *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Santos, Higuita. 2012. *Tingkat Kesukaran dan Daya Beda Butir-butir Soal Ujian Akhir Semester (UAS) Bahasa Indonesia di SMA Negeri 1 Batu Tahun Ajaran 2011/2012*. Artikel Skripsi. Malang: Fakultas Sastra Universitas Negeri Malang.
- Sastrawati, Sry Ihsa. 2005. *Analisis Butir Soal Objektif Mata Pelajaran Bahasa dan Sastra Indonesia dalam Tes Akhir Semester II Kelas 1 SMA Negeri 1 Nanga Pinoh Tahun Pembelajaran 2004/2005*. skripsi. Pontianak: FKIP UNTAN.
- Sudijono, Anas. 2007. *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada.

Sudjana, Nana. 2004. *Penilaian Hasil dan Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya

Suryabrata, Sumadi. 1987. *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali Press.

Wahidmuri, dkk. 2010. *Evaluasi Pembelajaran kompetensi dan praktik*. Yogyakarta: Nuha Litara

Widodo. 2010. *Analisis Butir Soal Tes*. Jurnal Pendidikan Penabur no 14/Tahun ke-9/Juni 2010.