

ANALISIS BUTIR SOAL FISIKA ULANGAN UMUM KENAIKAN KELAS X MADRASAH ALIYAH SE-KABUPATEN BANJARNEGARA, JAWA TENGAH TAHUN PELAJARAN 2011/2012

Lili Maenani^{a,b}, Raden Oktova^{b,c}

^{a)} MA Negeri 2 Banjarnegara, Jl. Letjend. Suprpto No. 95 A Kab. Banjarnegara
E-mail: lilimaenani@yahoo.co.id

^{b)} Program Magister Pendidikan Fisika, Program Pascasarjana
Universitas Ahmad Dahlan Yogyakarta
Kampus II, Jl. Pramuka 42 Lt.3, Yogyakarta 55161

^{c)} Program Studi Pendidikan Fisika, FKIP, Universitas Ahmad Dahlan Yogyakarta
Kampus III, Jl. Prof. Dr. Soepomo, Janturan, Yogyakarta 55166
E-mail: R.Oktova@uad.ac.id

INTISARI

Telah dilakukan penelitian deskriptif analisis butir soal Ulangan Umum Kenaikan Kelas X mata pelajaran Fisika pada Madrasah Aliyah (MA) se-Kabupaten Banjarnegara, Jawa Tengah, Tahun Pelajaran 2011/2012, dengan populasi penelitian sebanyak 980 siswa. Sebagai sampel penelitian diambil 277 siswa yang terdiri atas 178 siswa MA Negeri dan 99 siswa MA Swasta dengan teknik *proportionate stratified random sampling*. Butir soal yang dianalisis terdiri atas 35 butir soal pilihan ganda. Analisis dilakukan dengan menggunakan program Anates versi 4.0.2 untuk menentukan reliabilitas, daya pembeda, tingkat kesukaran dan validitas butir soal. Analisis butir soal menghasilkan koefisien reliabilitas tes 0,79, yang berarti soal memenuhi kriteria reliabilitas tinggi. Berdasarkan persentase daya pembeda diketahui sebagian besar butir soal memiliki daya pembeda sedang. Hasil perhitungan tingkat kesukaran soal menunjukkan bahwa lebih dari 50 % butir soal memiliki tingkat kesukaran sedang. Hasil perhitungan validitas butir soal menunjukkan 28,57% butir soal tidak valid, 28,57% valid dan 42,86% butir soal sangat valid.

Kata kunci: validitas, daya pembeda, reliabilitas, tingkat kesukaran, Ulangan Umum Kenaikan Kelas.

I. PENDAHULUAN

Bruce, Weil dan Calhoun [1] menyatakan bahwa pembelajaran pada hakekatnya merupakan suatu proses yang kompleks (rumit), namun dengan maksud yang sama yaitu memberikan pengalaman belajar kepada siswa sesuai dengan tujuan. Tujuan yang hendak dicapai sebenarnya merupakan acuan dalam penyelenggaraan proses pembelajaran. Untuk mengetahui ketercapaian tujuan pembelajaran maka perlu dilakukan evaluasi hasil belajar. Menurut Tyler [2], evaluasi adalah proses penentuan sejauh mana tujuan pendidikan telah tercapai. Definisi yang lebih luas dikemukakan oleh dua orang ahli lain, yakni Cronbach dan Stufflebeam [3-5], tambahan definisi tersebut adalah bahwa proses evaluasi bukan sekedar mengukur sejauh mana tujuan tercapai, tetapi digunakan untuk membuat keputusan.

Berdasarkan pengertian evaluasi di atas maka hakikat evaluasi pembelajaran adalah proses yang sistematis yaitu mengumpulkan data, menganalisis dan selanjutnya menarik kesimpulan tentang pencapaian hasil dan tingkat efektivitas dan efisiensi suatu program pembelajaran. Jadi dalam kegiatan evaluasi hasil belajar tidak terlepas dari kegiatan mengumpulkan data, sehingga diperlukan alat untuk mengumpulkan data yang disebut alat evaluasi atau instrumen. [6] menyebutkan terdapat dua cara untuk memperoleh data yaitu tes dan nontes. Dalam dunia pendidikan, dilihat dari bentuknya tes terdiri atas : tes tertulis (pilihan ganda, uraian, menjodohkan, jawaban singkat, mengisi, dan benar salah), tes lisan dan tes praktek. Yang tergolong teknik nontes adalah observasi (biasanya menggunakan alat berupa daftar cek, dan skala penilaian), angket/kuesioner, dan dokumentasi.

Alat evaluasi (soal ulangan atau tes) dapat dibuat langsung oleh guru pengajar maupun oleh tim yang ditunjuk. Khusus di lingkungan Madrasah Aliyah (MA) se-Karesidenan Banyumas, Jawa Tengah, soal Ulangan Akhir Semester (UAS) dan soal Ulangan Umum Kenaikan Kelas (UUKK), termasuk soal Fisika, disusun oleh sebuah tim, dan soal yang disusun digunakan oleh seluruh MA negeri maupun swasta se-Karesidenan Banyumas. Untuk mengetahui kualitas butir soal Fisika Ulangan Umum Kenaikan Kelas X Tahun pelajaran 2011/2012 perlu dilakukan analisis butir soal. Karena keterbatasan peneliti maka penelitian hanya dilakukan untuk wilayah Kabupaten Banjarnegara.

Berdasarkan latar belakang di atas, permasalahan yang diteliti adalah bagaimanakah reliabilitas, daya pembeda, tingkat kesukaran, dan validitas butir soal Fisika Ulangan Umum Kenaikan Kelas X MA se-Kabupaten Banjarnegara Tahun Pelajaran 2011/2012?

II. KAJIAN PUSTAKA

a. Evaluasi Hasil Belajar

Evaluasi pembelajaran merupakan proses yang sistematis dan terprogram sehingga harus dipersiapkan dengan sebaik-baiknya. Evaluasi pembelajaran dapat dikategorikan menjadi dua yaitu formatif dan sumatif. Evaluasi formatif adalah evaluasi yang dilakukan pada setiap pembahasan suatu pokok bahasan/topik. Winkel menyatakan bahwa yang dimaksud dengan evaluasi formatif adalah penggunaan tes-tes selama proses pembelajaran masih berlangsung, agar siswa dan guru memperoleh informasi mengenai kemajuan yang telah dicapai. Jadi evaluasi formatif bertujuan untuk memperbaiki proses belajar mengajar. Contoh evaluasi formatif adalah kuis, tugas dan ulangan harian. Adapun evaluasi sumatif adalah evaluasi yang dilakukan pada akhir setiap satu satuan waktu yang didalamnya tercakup lebih dari satu pokok bahasan. Winkel mendefinisikan evaluasi sumatif sebagai penggunaan tes-tes pada akhir suatu periode pengajaran tertentu, yang meliputi beberapa atau semua unit pelajaran yang diajarkan dalam satu semester, bahkan setelah selesai pembahasan suatu bidang studi. Evaluasi sumatif bertujuan untuk menetapkan tingkat keberhasilan peserta didik dalam kurun waktu tertentu [2]. Contoh dari evaluasi sumatif adalah Ulangan Tengah Semester (UTS), Ulangan Akhir Semester (UAS) pada akhir semester gasal dan Ulangan Umum Kenaikan Kelas (UUKK) pada akhir tahun pelajaran.

Agar hasil evaluasi dapat digunakan dengan baik maka alat evaluasi harus valid dan reliabel. Evaluasi hasil belajar bertujuan untuk mengetahui ketercapaian hasil belajar siswa jadi instrumen evaluasi harus mampu memotret kemampuan sesungguhnya yang dimiliki siswa dengan kata lain alat evaluasi tes harus valid [7]. Secara garis besar ada dua macam validitas, yaitu validitas logis (*logical validity*) dan validitas empiris (*empirical validity*) [3]. Sebuah instrumen dinyatakan memiliki validitas logis jika memenuhi persyaratan valid berdasarkan hasil penalaran, yaitu dirancang dengan baik mengikuti teori dan ketentuan yang sudah ada. Dengan demikian validitas logis tidak perlu diuji kondisinya tetapi langsung diperoleh sesudah instrumen tersebut selesai disusun. Validitas logis dibagi menjadi dua, yaitu validitas isi dan validitas konstruk. Validitas isi menyangkut instrumen yang disusun berdasarkan isi materi pelajaran yang dievaluasi sedangkan validitas konstruk menunjuk pada suatu kondisi sebuah instrumen yang disusun berdasarkan konstruk aspek-aspek kejiwaan yang harus dievaluasi. Sebuah instrumen dikatakan memiliki validitas empiris apabila sudah diuji dari pengalaman. Pengujian dilakukan dengan membandingkan kondisi instrumen dengan kriterium atau sebuah ukuran. Instrumen yang kondisinya sudah sesuai dengan kriterium yang sudah ada disebut memiliki validitas yang ada sekarang (*concurrent validity*), sedangkan instrumen yang kondisinya sesuai dengan kriterium yang diramalkan akan terjadi dikatakan memiliki validitas ramalan (*predictive validity*).

Stainback [7] menyatakan bahwa reliabilitas berkenaan dengan derajad konsistensi dan stabilitas data atau temuan. Dalam pandangan positivistik (kuantitatif), suatu data dinyatakan reliabel apabila dua atau lebih peneliti dalam obyek yang sama menghasilkan data yang sama, atau peneliti sama dalam waktu berbeda menghasilkan data yang sama, atau sekelompok data bila dipecah menjadi dua menunjukkan data yang tidak berbeda. Reliabel dapat diartikan sebagai 'ajeg' atau 'tetap' sehingga instrumen yang reliabel dapat mengungkap keadaan suatu objek secara ajeg meskipun digunakan pada waktu yang berbeda.

b. Analisis Butir Soal

Validitas dan reliabilitas soal atau instrumen dipengaruhi oleh kualitas tiap butir soal, sehingga perlu dilakukan analisis butir soal. Analisis butir soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir soal [3]. Analisis butir soal pada dasarnya terbagi menjadi dua kategori, yaitu analisis butir soal secara kualitatif dan secara kuantitatif [2]. Analisis secara kualitatif lebih menekankan pada aspek perakitan soal sehingga dapat dilakukan tanpa ujicoba, yaitu telaah yang dilakukan terhadap aspek materi, aspek konstruksi dan aspek bahasa. Adapun analisis kuantitatif menekankan pada karakteristik internal tes melalui data yang diperoleh secara empiris. Karakteristik internal butir soal yang

dimaksud meliputi parameter daya validitas, reliabilitas, pembeda atau indeks diskriminasi, tingkat kesukaran dan pola jawaban soal.

III. METODE PENELITIAN

a. Desain Penelitian

Penelitian yang dilakukan adalah penelitian deskriptif yaitu penelitian yang bertujuan untuk menggambarkan suatu variabel mandiri baik hanya satu variabel atau lebih [5]. Jadi peneliti tidak membuat perbandingan variabel tersebut pada sampel yang lain, atau mencari hubungan antar variabel. Penelitian dilakukan untuk mengetahui kualitas butir soal Fisika Ulangan Umum Kenaikan Kelas Kelas X MA se-Kabupaten Banjarnegara Tahun Pelajaran 2011/2012.

b. Populasi dan sampel

Populasi penelitian adalah 980 siswa kelas X MA se-Kabupaten Banjarnegara Tahun Pelajaran 2011/2012. Berdasarkan informasi dari Kantor Kementerian Agama Kabupaten Banjarnegara Seksi MAPENDA (Madrasah Pendidikan Agama) Islam, Tabel I menunjukkan jumlah siswa kelas X MA se-Kabupaten Banjarnegara Tahun Pelajaran 2011/2012.

Tabel I. Jumlah siswa kelas X MA se-Kabupaten Banjarnegara

No.	MA	Jumlah Siswa	Persentase
1.	Negeri	629	64,2 %
2.	Swasta	351	35,8 %
	Jumlah	980	100 %

Untuk menentukan jumlah sampel digunakan rumus dari Isaac dan Michael [7]

$$S = \frac{\}^2 NPQ}{d^2(N-1) + \}^2 PQ}, \quad (1)$$

dengan $\}^2 = 3,841$ dk = 1, taraf kesalahan bias 5 % $P = Q = 0,5$, $d = 0,05$, dan $S =$ ukuran sampel. Dengan rumus di atas jumlah sampel penelitian adalah 277 siswa. Teknik pengambilan sampel yang digunakan adalah teknik *proportionate stratified random sampling* [7]. Teknik ini digunakan karena populasi mempunyai anggota yang tidak sama jumlahnya dari tiap-tiap madrasah. Sebagaimana dapat dilihat pada Tabel I, sampel penelitian terdiri atas 178 orang atau 64,2% siswa MA negeri dan 99 orang atau 35,8% siswa MA swasta.

c. Variabel Penelitian

Variabel yang diteliti adalah reliabilitas, daya pembeda, tingkat kesukaran dan validitas butir soal Fisika Ulangan Umum Kenaikan Kelas X MA se-Kabupaten Banjarnegara Tahun Pelajaran 2011/2012.

d. Uji Validitas

Teknik untuk mengetahui validitas butir soal adalah dengan menentukan kesejajaran antara dua variabel menggunakan teknik korelasi *product moment* yang dikemukakan oleh Pearson [6-9],

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

dengan r_{XY} = koefisien korelasi antar variabel X dan variabel y , n = cacah responden, Y = skor total yang diperoleh responden dari semua butir soal dan X = skor butir soal yang diperoleh responden. Untuk mengetahui apakah setiap butir dalam instrumen tersebut valid atau tidak, dapat diketahui dengan cara mengkorelasikan antara skor butir (X) dengan skor total (Y) atau biasanya disebut validitas item. Bila nilai korelasi dibawah 0,30, dapat disimpulkan bahwa butir instrumen tersebut tidak valid, sehingga harus diperbaiki atau dibuang [5].

e. Uji Reliabilitas

Uji reliabilitas bertujuan untuk mengetahui konsistensi hasil pengukuran dari instrumen tersebut. Untuk uji reliabilitas digunakan rumus Kuder Richardson-20 (KR-20) [10,11]

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \sum pq}{S^2} \right), \quad (3)$$

dengan r_{11} reliabilitas tes secara keseluruhan, p proporsi sebyek yang menjawab benar, q proporsi subyek yang menjawab salah, $\sum pq$ jumlah hasil perkalian antara p dan q , dan S deviasi standard. Kriteria koefisien reliabilitas menurut Arikunto [3] dapat dilihat pada Tabel II.

Tabel II. Kriteria tingkat reliabilitas.

Koefisien reliabilitas	Tingkat reliabilitas
$0,00 < r_{11} \leq 0,20$	sangat rendah
$0,20 < r_{11} \leq 0,40$	Rendah
$0,40 < r_{11} \leq 0,60$	Sedang
$0,60 < r_{11} \leq 0,80$	Tinggi
$0,80 < r_{11} \leq 1,00$	Sangat tinggi

f. Uji Daya Pembeda

Soal yang baik dapat membedakan antara siswa yang pandai dan yang kurang pandai. Kemampuan soal untuk membedakan antara siswa yang pandai dengan siswa yang berkemampuan kurang disebut daya pembeda [3]. Soal yang memiliki daya pembeda tinggi jika diberikan pada sekelompok siswa yang pandai hasilnya tinggi sebaliknya jika diberikan pada siswa yang kurang pandai maka hasilnya rendah.

Untuk menentukan indeks diskriminasi (daya pembeda) digunakan rumus

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B, \quad (4)$$

dengan D = indeks diskriminasi, J_A = banyaknya peserta kelompok atas, J_B = banyaknya peserta kelompok bawah, B_A = banyaknya peserta kelompok atas yang menjawab benar, B_B = banyaknya peserta kelompok bawah yang menjawab benar, P_A = proporsi peserta kelompok atas yang menjawab benar, P_B = proporsi peserta kelompok bawah yang menjawab benar [3]. Indeks diskriminasi (daya pembeda) berkisar antara -1 sampai +1,00. Pada indeks diskriminasi kadang menghasilkan nilai negatif hal ini berarti untuk item soal tersebut siswa pada kelompok bawah lebih banyak yang menjawab benar dibandingkan dengan siswa pada kelompok atas yang menjawab benar. Dengan demikian soal dengan daya pembeda negative harus harus dibuang. Crocker dan Algina [2] membuat kriteria besar koefisien daya pembeda soal seperti ditunjukkan Tabel III.

Tabel III. Kriteria daya pembeda

No	Koefisien korelasi	Kategori
1.	-1,00 sampai dengan 0,19	Tidak berfungsi
2.	0,20 sampai dengan 0,29	Perlu revisi
3.	0,30 sampai dengan 0,39	Tidak perlu revisi
4.	0,40 sampai dengan 1,00	Baik

Arikunto [3] memberikan klasifikasi daya pembeda seperti pada Tabel IV.

Tabel IV. Klasifikasi daya pembeda

No	Koefisien korelasi	Kategori
1.	0,00 sampai dengan 0,20	Jelek
2.	0,20 sampai dengan 0,40	Cukup
3.	0,40 sampai dengan 0,70	Baik
4.	0,70 sampai dengan 1,00	Baik sekali

Dengan demikian soal dengan kategori daya pembeda jelek adalah soal yang tidak berfungsi dan soal dengan kategori baik dan sangat baik sudah dapat diterima tanpa revisi, sedangkan soal dengan kategori cukup ada beberapa yang perlu direvisi yaitu butir soal yang memiliki koefisien daya pembeda antara 0,20 sampai 0,29.

g. Tingkat kesukaran

Soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar atau memiliki tingkat kesulitan dengan proporsi tertentu. Tingkat kesukaran adalah bilangan yang menunjukkan sukar atau mudahnya suatu soal, dan disebut juga tingkat kesukaran [3].

Untuk mencari tingkat kesukaran, P digunakan rumus

$$P = \frac{B}{J_s}, \quad (5)$$

dengan P = tingkat kesukaran, B = banyaknya siswa yang menjawab benar, dan J_s = cacah seluruh siswa peserta tes [3]. Tingkat kesukaran memiliki nilai 0,00 sampai dengan 1,0. Soal dengan tingkat kesukaran 0,00 menunjukkan bahwa soal tersebut terlalu sukar, sebaliknya soal dengan tingkat kesukaran 1,0 menunjukkan bahwa soal tersebut terlalu mudah. Adapun kriteria tingkat kesukaran [3] dapat dilihat pada Tabel V.

Tabel V. Kriteria tingkat kesukaran soal

No	Tingkat kesukaran	Kategori
1.	1,00 sampai dengan 0,30	Sukar
2.	0,30 sampai dengan 0,70	Sedang
3.	0,70 sampai dengan 0,00	Mudah

Alat evaluasi disusun dengan tujuan tertentu jika tujuannya untuk seleksi maka digunakan soal-soal dengan tingkat kesukaran yang tinggi, dengan demikian hanya peserta tes dengan kemampuan tinggi saja yang

dapat mengerjakan. Soal tes yang digunakan untuk mengetahui hasil belajar sebaiknya memiliki komposisi tingkat kesukaran tertentu. Rasyid dan Mansur [2] menyatakan proporsi jumlah soal untuk kategori mudah, sedang dan sukar dapat didasarkan atas kurva normal yaitu dengan komposisi 25 % berkategori mudah, 50 % berkategori sedang dan 25 % lainnya berkategori sukar.

h. Teknik Pengumpulan Data

Data penelitian ini adalah jawaban responden yang diperoleh dari lembar jawab peserta Ulangan Umum Kenaikan Kelas X MA se-Kabupaten Banjarnegara. Berdasarkan penentuan sampel di atas maka diambil secara acak lembar jawab siswa pada tiap-tiap MA.

i. Teknik Analisis Data

Teknik analisis data yang digunakan adalah statistik deskriptif yaitu statistik yang mempunyai tugas mengorganisasi dan menganalisis data angka agar dapat memberikan gambaran secara teratur, ringkas, dan jelas mengenai suatu gejala, peristiwa atau keadaan, sehingga dapat ditarik pengertian atau makna tertentu [11]. Statistik deskriptif untuk menganalisis butir soal dapat dilakukan dengan menggunakan program Anates versi 4.0.2 yang dikembangkan oleh Karno To dan Yudi Wibisono [12-13].

IV. HASIL DAN PEMBAHASAN

a. Reliabilitas

Reliabilitas merupakan keajegan soal tes untuk mengungkap prestasi/hasil belajar siswa. Hasil perhitungan korelasi XY diperoleh nilai $r = 0,79$. Berdasarkan Tabel II dapat diinterpretasikan bahwa soal Fisika Ulangan Umum Kenaikan Kelas X MA se-Kabupaten Banjarnegara memiliki reliabilitas yang tinggi.

b. Daya Pembeda

Berdasarkan Tabel III tentang kategori daya pembeda menurut Croker dan Algina, 5 butir soal (14,28%) tidak berfungsi yaitu nomor 10, 15, 27, 28, dan 35, bahkan butir soal nomor 10 harus dibuang karena memiliki koefisien daya pembeda bernilai negatif. Terdapat 4 butir soal (11,43%) yang perlu direvisi yaitu butir soal nomor 1, 21, 23, dan 34. Adapun 26 (74,29%) butir soal lainnya tidak perlu direvisi.

Adapun berdasarkan kriteria Arikunto (Tabel IV) tentang klasifikasi daya pembeda diketahui 5 butir soal atau 14,28% butir soal memiliki daya pembeda yang jelek yaitu butir soal nomor 10, 15, 27, 28, 35. Sebelas butir soal atau 31,43% butir soal memiliki daya pembeda cukup yaitu soal nomor 1, 2, 5, 6, 8, 14, 21, 23, 26, 31, 34. Sembilan belas butir soal atau 54,29% butir soal memiliki daya pembeda baik, yaitu butir soal nomor 3, 4, 7, 9, 11, 12, 13, 16, 17, 18, 19, 20, 22, 24, 25, 29, 30, 32, dan 33. Dengan demikian dapat disimpulkan bahwa sebagian besar butir soal memiliki daya pembeda sedang.

Sesungguhnya untuk meningkatkan daya pembeda, dapat dilakukan uji coba soal sebelum digunakan, namun soal Ulangan Umum Kenaikan Kelas dalam kenyataannya tidak diujicobakan karena dikhawatirkan terjadi kebocoran. Salah satu alternatif solusi untuk memperoleh soal dengan daya pembeda baik adalah dengan mengoptimalkan proses penyusunan soal sesuai kaidah, terutama pada perakitan butir soal. Rumusan soal harus jelas sehingga tidak menimbulkan penafsiran ganda. Gambar, grafik, tabel dan diagram yang terdapat dalam soal harus jelas dan berfungsi. Dari segi bahasa, soal harus menggunakan bahasa yang komunikatif dan sesuai dengan kaidah bahasa Indonesia, serta tidak menggunakan bahasa yang berlaku setempat. Pilihan jawaban (*option*) juga sangat mempengaruhi daya pembeda soal. Pilihan jawaban harus homogen dan logis dengan materi, artinya tidak ada pilihan jawaban yang terlalu menyimpang dari materi. Panjang rumusan pilihan jawaban juga harus relatif sama, karena ada kecenderungan siswa memilih jawaban yang paling panjang. Distraktor atau pengecoh harus berfungsi, jadi siswa yang tidak memahami materi tes dimungkinkan memilih pengecoh.

c. Tingkat Kesukaran

Hasil perhitungan tingkat kesukaran soal setelah dikonsultasikan dengan Tabel V tentang kriteria tingkat kesukaran soal, menunjukkan bahwa 3 butir soal atau 8,57% dikategorikan mudah yaitu butir soal nomor 3, 21, 31. Dua puluh butir soal atau 57,14% memiliki tingkat kesukaran dengan kategori sedang, yaitu butir soal nomor 2, 4, 5, 9, 11, 12, 13, 14, 16, 17, 18, 20, 22, 25, 26, 30, 32, 33, 34, 35. Sebelas butir soal atau 31,43% dikategorikan sukar yaitu butir soal nomor 1, 6, 7, 8, 10, 15, 19, 23, 24, 27 dan 29 dan 1 butir soal atau 2,86% dikategorikan sangat sukar yaitu butir soal nomor 28. Jadi sebagian besar (lebih dari 50 %) butir soal Fisika Ulangan Umum Kenaikan Kelas X MA se-Kabupaten Banjarnegara memiliki tingkat kesukaran sedang.

Berdasarkan hasil analisis tingkat kesukaran, soal yang diteliti belum mengikuti komposisi yang ideal yaitu 25% mudah, 50% sedang dan 25% sukar. Salah satu cara untuk memperoleh komposisi soal dengan kriteria tingkat kesukaran yang sesuai adalah dengan meminta penilaian (*judgment*) dari guru yang sudah berpengalaman. Meskipun perkiraan dari guru ini belum tentu sepenuhnya benar (misalnya, mudah menurut guru tersebut belum tentu mudah bagi siswa peserta tes), setidaknya ada usaha untuk memperoleh soal dengan tingkat kesukaran yang baik.

d. Validitas

Hasil perhitungan untuk mengetahui validitas butir soal menghasilkan koefisien korelasi tiap butir soal. Berdasarkan teori di atas, jika koefisien korelasi kurang dari 0,30 maka butir soal tersebut tidak signifikan atau tidak valid, sebaliknya jika koefisien korelasi lebih dari 0,30 maka butir soal tersebut signifikan atau valid. Berdasarkan hasil perhitungan diperoleh sepuluh butir soal atau 28,57% butir soal tidak signifikan, yaitu butir soal nomor 1, 2, 10, 15, 21, 27, 28, 31, 34, dan 35. Sepuluh butir soal atau 28,57% signifikan yaitu soal nomor 3, 5, 12, 13, 14, 18, 23, 26, 30, dan 33. Lima belas butir soal atau 42,86% butir soal sangat signifikan yaitu soal nomor 4, 6, 7, 8, 9, 11, 16, 17, 19, 20, 22, 24, 25, 29, dan 32. Secara singkat dapat disampaikan bahwa 28,57% butir soal tidak valid, 28,57% valid dan 42,86% butir soal sangat valid.

Perlu ditekankan bahwa uji validitas yang dilakukan dalam penelitian ini adalah uji validitas empiris saja. Cara lain untuk menguji validitas adalah uji validitas logis yang dilakukan tanpa uji coba, yaitu dengan merancang soal sesuai dengan teknik pembuatan soal yang baik. Kisi-kisi soal dibuat terlebih dahulu sebelum soal dirakit sehingga substansi soal sesuai dengan materi yang diajarkan.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan, maka dapat disimpulkan bahwa soal Fisika Ulangan Umum Kenaikan Kelas X MA se-Kabupaten Banjarnegara Tahun Pelajaran 2011/2012 (i) reliabilitas tinggi, (ii) sebagian besar butir soal memiliki daya pembeda sedang (iii) sebagian besar butir soal memiliki tingkat kesukaran sedang (iv) 28,57% butir soal tidak valid, 28,57% valid dan 42,86% butir soal sangat valid.

Keterbatasan penelitian ini adalah telaah butir soal yang dilakukan hanya telaah secara kuantitatif (empiris), yaitu telaah dilakukan setelah pelaksanaan Ulangan Umum Kenaikan Kelas, sehingga untuk penelitian selanjutnya perlu dilakukan telaah secara kualitatif (teoretis), yaitu telaah yang dilakukan terhadap aspek materi, aspek konstruksi dan aspek bahasa.

DAFTAR PUSTAKA

- [1] Sumiati dan Asra, 2008, "Metode Pembelajaran", Bandung: Wacana Prima.
- [2] Rasyid, H., dan Mansur, 2008, "Penilaian Hasil Belajar", Bandung: Wacana Prima.
- [3] Arikunto, S., 2009, "Dasar-dasar Evaluasi Pendidikan", Jakarta: Bumi Aksara.
- [4] Cronbach, L., 2000, "Course Improvement through Evaluation", dalam D. I. Stufflebeam, G. F. Madoux, and T. Kellaghan (Eds.), *Evaluation Models* (pp. 16–32). Boston, MA: Kluwer Academic Publishers.
- [5] Stufflebeam, D.L., 2001, "Evaluation Models," *New Directions for Evaluation*, **89**, 7–98.
- [6] Tola, B., dan Fahmi, 2003, "Standar Penilaian di Kelas", Jakarta: Depag RI.
- [7] Sugiyono, 2010, "Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D", Bandung: Alfabeta.
- [8] Rodgers, J. L., and Nicewander, W. A., 1988, "Thirteen Ways to Look at the Correlation Coefficient," *The American Statistician*, **42**(1), 59–66.
- [9] Stigler, S.M., 1989, "Francis Galton's Account of the Invention of Correlation," *Statistical Science*, **4**(2), 73–79.
- [10] Kuder, G.F., and Richardson, M.W., 1937, "The Theory of the Estimation of Test Reliability," *Psychometrika*, **2**(3), 151–160.
- [11] Cortina, J. M., 1993, "What is Coefficient Alpha? An Examination of Theory and Applications," *Journal of Applied Psychology*, **78**(1), 98–104.
- [12] Sudijono, A., 2009, "Pengantar Statistika Pendidikan", Jakarta: Rajawali Press.
- [13] Anonim, 2010, "Manfaat ANATES 4: Analisis Tes Pilihan Ganda", <http://mydiaryzone.blogspot.com/2010/12/manfaat-anates-4-analisis-tes-pilihan.html>, 2010, diakses 14 September 2012.