

**ANALISIS INSTRUMEN TES ULANGAN AKHIR SEMESTER GENAP  
MATA PELAJARAN FISIKA SMA KELAS XI DI KABUPATEN  
BOYOLALI TAHUN PELAJARAN 2012/2013**

Triyani<sup>1</sup>, Nonoh Siti Aminah<sup>2</sup>, Surantoro<sup>2</sup>

Fakultas Keguruan dan Ilmu Pendidikan

Universitas Sebelas Maret

Jl. Ir. Sutami 36 A, Surakarta, email : 3ynee.byl@gmail.com

**ABSTRACT**

*This study aims to explain the characteristics of the test both qualitatively and quantitatively about the end of Deuteronomy (UAS) Even subjects in high school physics class XI Boyolali school year 2012/2013 and determine the effect of concentration on the amount of data measurements with classical test theory.*

*Type of research is the study of measurement. The sample was composed of a package about the matter, the student response data in 1036 and a number of the key answers to the end of (UAS) Even subjects in high school physics class XI Boyolali school year 2012/1013. Data obtained by the techniques of documentation. Data made into two groups of data is data 1 and data 2. Data 1 and Data 2 were analyzed qualitatively and quantitatively. Qualitative analysis with expert judgment technique. Quantitative analysis was performed using a computer-assisted program MicroCat ITEMAN version 3.00 and SPSS.*

*The results of the research that has been done is ( 1 ) about the qualitative characteristics contained 18 items about 45 % have met all three aspects of the study and 22 grains of about 55 % less than the three aspects of the study were met, about the reliability quantitatively quite good for the natural sciences; about the difficulty index is classified, Mean p on the results of the data ITEMAN 1 shows the numbers 0,526 and 0,681 figure 2 shows the data ; different power each item on the otherwise well , Biser point value in the results ITEMAN nothing negative. Additionally ITEMAN program printout indicates the number of data biserial Mean 1 ( 0.542 ) and data 2 ( 0.374 ) ; pattern of response to function properly known problem of the data contained 138 1 86.25 % effective functioning distractors and distractor 109 data 2 are 68.125 % effective. ( 2 ) The effect of data centralization to scale parameter measurements with classical test theory of a data 1 to data 2 is the reliability value of 0.879 into 0.678, the mean p value of 0.526 into 0.681, the mean biserial value of 0.542 into 0.374 , the mean value of 21.035 into 27.230 , the median of 21 to 27 , mode of 15 to 28, variance of 61 , 310 becomes 21.706 , standard deviation shift of 7.830 into 4.659 .*

*Keywords : expert judgment , central tendency , ITEMAN , the grain problem*

**ABSTRAK**

Penelitian ini bertujuan menjelaskan karakteristik tes baik secara kualitatif maupun kuantitatif soal Ulangan Akhir Semester (UAS) Genap mata pelajaran Fisika kelas XI SMA di Kabupaten Boyolali tahun pelajaran 2012/2013 dan mengetahui efek pemusatan data pada besaran pengukuran dengan teori tes klasik.

Jenis penelitian yang digunakan adalah penelitian pengukuran. Sampel penelitian ini adalah paket soal yang terdiri dari soal, data jawaban siswa sejumlah 1036 dan kunci jawaban soal Ulangan Akhir Semester (UAS) Genap mata pelajaran Fisika kelas XI SMA di Kabupaten Boyolali tahun pelajaran 2012/1013. Data diperoleh dengan teknik dokumentasi. Data dikelompokkan menjadi dua kelompok yaitu data 1 dan data 2. Data 1 dan data 2 dianalisis secara kualitatif dan kuantitatif. Analisis kualitatif dengan teknik *expert judgement*. Analisis kuantitatif dilakukan dengan bantuan komputer menggunakan program *MicroCat ITEMAN versi 3.00* dan program SPSS.

Hasil penelitian yang telah dilakukan adalah (1) karakteristik soal secara kualitatif terdapat 18 butir soal 45 % telah memenuhi ketiga aspek telaah dan 22 butir soal 55 % kurang dari tiga aspek telaah yang terpenuhi, secara kuantitatif reliabilitas soal tergolong baik untuk ilmu alam; indeks kesukaran soal tergolong sedang, Mean p pada hasil ITEMAN data 1 menunjukkan angka 0,526 dan data 2 menunjukkan angka 0,681; daya beda setiap butir soal dinyatakan baik, Nilai point biser dalam hasil ITEMAN tidak ada yang bernilai negatif. Selain itu hasil print out program ITEMAN menunjukkan angka Mean biserial data 1 (0,542) dan data 2 (0,374); pola jawaban soal berfungsi dengan baik diketahui dari data 1 terdapat 138 distraktor 86,25 % berfungsi efektif dan data 2 terdapat 109 distraktor 68,125 % berfungsi efektif. (2) Efek pemusatan data terhadap parameter besaran pengukuran dengan teori tes klasik dari data 1 ke data 2 yaitu nilai reliabilitas dari 0,879 menjadi 0,678, nilai mean p dari 0,526 menjadi 0,681, nilai mean biserial dari 0,542 menjadi 0,374, nilai mean dari 21,035 menjadi 27,230, median dari 21 menjadi 27, modus dari 15 menjadi 28. varians dari 61, 310 menjadi 21,706, simpangan baku bergeser dari 7,830 menjadi 4,659.

Kata kunci : *expert judgement, tendency central, ITEMAN, butir soal*

## 1. PENDAHULUAN

Tugas guru dalam mewujudkan tujuan pembelajaran, salah satunya dengan mengadakan evaluasi. Evaluasi pembelajaran dapat memberikan informasi mengenai perkembangan pemahaman siswa-siswanya tentang tingkat keberhasilan proses pembelajaran yang telah dilaksanakan.

Ditinjau dari cakupannya evaluasi ada dua yaitu evaluasi yang bersifat makro dan evaluasi yang bersifat mikro (Aminah, 2012: 1). Salah satu evaluasi yang bersifat mikro adalah Ulangan Akhir Semester (UAS). Pada penelitian yang dilakukan menggunakan evaluasi yang bersifat mikro yaitu Ulangan Akhir Semester (UAS) Mata Pelajaran Fisika Kelas XI SMA di Boyolali.

Peneliti menggunakan teori tes klasik untuk mengetahui karakteristik tes. Aspek yang perlu diperhatikan dalam analisis butir soal secara klasik adalah dari segi validitas, reliabilitas tes, tingkat kesukaran butir soal, daya beda, dan penyebaran pilihan jawaban (Depdiknas, 2008: 8). Salah satu program komputer yang dapat membantu kegiatan menganalisis butir soal yaitu *MicroCat ITEMAN version 3.00*

## 2. Tinjauan Pustaka

Tinjauan pustaka ini akan menguraikan tentang pengukuran, penilaian dan evaluasi; tujuan dan fungsi evaluasi; teknik evaluasi; alat evaluasi; kualitas soal; analisis

butir soal (*item analysis*) dan ukuran pemusatan data (*central tendency*).

#### a. Pengukuran, Penilaian dan Evaluasi

Proses pengukuran, penilaian, evaluasi, dan pengujian merupakan suatu kegiatan atau proses yang bersifat hierarkis. Kegiatan dilakukan secara berurutan dan berjenjang, dimulai dari proses pengukuran kemudian penilaian dan terakhir evaluasi (Haryati, 2007: 14).

Mengukur, menurut Azwar (2002: 4) adalah membandingkan atribut yang hendak diukur dengan alat ukurnya secara deskriptif. Pengukuran adalah suatu proses kegiatan untuk memberikan angka. Angka di sini dimaknai sebagai skor yang merupakan langkah awal dalam proses evaluasi. Hal ini sejalan dengan pendapat Silverius (1991:6) menjelaskan pengukuran adalah suatu proses pemberian angka pada sesuatu atau seseorang berdasarkan aturan-aturan tertentu. Demikian juga dengan pendapat Widoyoko (2009: 2) yang menyebutkan bahwa esensi dari pengukuran adalah kuantifikasi atau penetapan angka tentang karakteristik atau keadaan individu menurut aturan-aturan tertentu.

Menilai, menurut Azwar (2002: 157) adalah proses penerjemahan skor hasil tes yang telah dikonversikan, ke dalam klasifikasi menurut norma atau kriteria yang relevan. Penilaian bersifat kualitatif. Menilai adalah memberikan nilai terhadap sesuatu yang merupakan proses menterjemahkan skor hasil tes. Nilai dapat berupa angka dan non angka. Berbeda dengan pengukuran yang hasilnya berupa angka-angka, penilaian memberikan nilai berupa non angka, seperti memuaskan atau tidak memuaskan, baik atau buruk, sehat atau sakit, pandai atau bodoh, tinggi atau rendah, dan sebagainya. Seperti halnya dengan pendapat Widoyoko (2009: 3) yang menyebutkan bahwa penilaian dapat diartikan sebagai kegiatan menafsirkan data hasil pengukuran berdasarkan kriteria maupun aturan-aturan tertentu.

Evaluasi, menurut Indrastoeti (2012: 4) evaluasi adalah proses pemberian makna atau penetapan kualitas hasil pengukuran dengan cara membandingkan angka hasil pengukuran tersebut dengan kriteria tertentu. Dimana evaluasi didahului penilaian sedangkan penilaian didahului dengan pengukuran.

Pengertian tentang pengukuran, penilaian, dan evaluasi di atas menunjukkan bahwa pengukuran, penilaian, dan evaluasi merupakan tiga konsep yang berbeda. Namun, dalam praktiknya di dunia pendidikan, ketiga konsep tersebut sering dipraktikkan dalam satu rangkaian.

#### b. Tujuan dan Fungsi Evaluasi

Evaluasi di bidang pendidikan sangat bermanfaat baik bagi siswa, guru, sekolah, maupun bagi supervisor. Kegiatan evaluasi harus dilaksanakan dengan teliti karena hasilnya menyangkut orang banyak dan berdampak luas. Hasil evaluasi dapat memberikan informasi yang akurat dan objektif tentang suatu program, sehingga berguna dalam pengambilan keputusan atau kebijakan suatu program.

#### c. Teknik Evaluasi

Dalam proses belajar mengajar di kelas ada banyak aspek yang harus diukur. Misalnya kemampuan yang berhubungan dengan pengetahuan, keterampilan siswa, minat, sikap, dan kepribadian. Untuk mengukur aspek-aspek tersebut digunakan teknik penilaian yang berbeda-beda.

Secara garis besar, teknik penilaian yang digunakan dapat digolongkan menjadi 2 macam, yaitu: tes dan bukan tes (nontes) (Suharsimi, 2006: 26). Teknik tes, yaitu teknik penilaian yang menggunakan tes sebagai alat ukurnya. Teknik non-tes, yaitu teknik penilaian yang menggunakan bukan tes sebagai alat ukurnya. Termasuk dalam kategori teknik non-testing adalah observasi/pengamatan yang dapat berbentuk rating scale, anecdotal record, atau rekaman, interview, questionnaire, dan inventori (Tim PEKERTI-AA PPSP LPP, 2007: 21).

Teknik penilaian umumnya digunakan untuk menilai kemampuan yang berhubungan dengan pengetahuan dan ketrampilan siswa, bakat khusus dan bakat umum. Sementara untuk teknik nontes umumnya digunakan untuk menilai karakteristik-karakteristik lain dari siswa misalnya minat, sikap dan kepribadian (Cartono Utari dan Toto Sutarto G, 2006: 45).

#### d. Alat Evaluasi

Tes, menurut Azwar (2002) dilihat dari wujud fisiknya, suatu tes tidak lain dari sekumpulan pertanyaan yang harus dijawab atau tugas yang harus dikerjakan yang akan memberikan informasi mengenai aspek psikologis tertentu berdasarkan jawaban terhadap pertanyaan-pertanyaan atau cara dan hasil subjek dalam melakukan tugas-tugas tersebut (hlm 2). Sama halnya dengan pendapat Aminah (2012) yang menyatakan tes adalah suatu alat atau metode pengumpul data yang sistematis untuk mengukur dan atau menilai salah satu aspek kemampuan atau kecakapan dengan jalan mengukur sampel dari salah satu aspek tersebut (hlm 2).

Jenis tes berdasarkan waktu penyelenggaraannya terdapat 3 jenis, yaitu tes masuk (*entrance test*), tes formatif (*formative test*) dan tes sumatif (*summative test*) (Indrastoeti, 2012: 109). Menurut Sudijono (2005) ada dua fungsi umum yang dimiliki oleh tes yaitu:

- a) Sebagai alat pengukur terhadap peserta didik. Dalam hubungan ini tes berfungsi mengukur tingkat perkembangan atau kemajuan yang telah dicapai oleh peserta didik setelah menempuh proses belajar mengajar dalam jangka waktu tertentu.
- b) Sebagai alat pengukur keberhasilan program pengajaran, sebab melalui tes tersebut akan dapat diketahui sudah seberapa jauh program pengajaran yang telah ditentukan, telah dapat dicapai (hlm 67).

Secara garis besar Silverius (1991: 18) menyebutkan bahwa, bentuk soal dibedakan menjadi dua macam, yaitu:

- 1) Bentuk objektif
- 2) Bentuk uraian/subjektif (*essay*)

Bentuk uraian atau *essay* itu dapat dibedakan lagi atas:

- a) Uraian bebas, disebut juga uraian *non-objektif*
- b) Uraian terbatas (*restricted*), disebut juga *uraian objektif*.

#### e. Kualitas Soal

Soal yang berkualitas harus memenuhi kriteria sebagai alat ukur yang baik. Kriteria alat ukur yang baik ditinjau dari teori tes klasik yaitu reliabilitas tes, indeks daya beda, indek kesukaran dan validitas tes (Aminah, 2012: 2). Suatu alat penilaian dikatakan mempunyai kualitas yang baik apabila alat tersebut memiliki atau memenuhi dua hal, yakni ketepatannya atau validitasnya dan keajegannya atau reliabilitasnya (Sudjana, 2008: 12). Menurut Purwanto (1988: 178), mengukur kesesuaian, efisiensi dan kemandapan (*consistency*) suatu alat penilaian atau suatu tes dipergunakan bermacam-macam kualitas seperti validitas, reliabilitas, objektivitas, dan kepraktisan (*practicibility*). Demikian, Widoyoko (2009: 98-102) menyebutkan bahwa sebuah tes yang dapat dikatakan baik sebagai alat pengukuran harus memiliki persyaratan tes, yaitu: validitas, reliabilitas, objektivitas dan ekonomis.

Pendapat para pakar di atas, dapat diambil kesimpulan bahwa kualitas soal yang baik atau berkualitas harus memiliki persyaratan tes berupa validitas, reliabilitas, indeks daya beda, indeks kesukaran, objektivitas, praktikabilitas dan ekonomis. Namun, syarat terpenting dalam menentukan kualitas soal adalah validitas dan reliabilitas.

#### f. Analisis Butir Soal (*Item Analysis*)

Analisis butir soal dilaksanakan dalam dua langkah, yakni analisis secara kuantitatif dan kualitatif. Analisis kualitatif

berhubungan dengan isi dan bentuk butir soal sedangkan analisis kuantitatif berhubungan dengan ciri statistiknya. Ciri statistik yang dimaksudkan antara lain reliabilitas, daya beda, tingkat kesukaran, dan efektifitas pengecoh.

**Analisis butir soal secara kualitatif**

Analisis butir soal secara kualitatif merupakan telaah soal dari segi materi, konstruksi, dan aspek materi yang dilakukan oleh ahli dalam bidang-bidang tersebut. Dalam melakukan analisis kualitatif pemilihan format penelaah soal sangat membantu mempermudah pekerjaan analisis soal (Depdiknas, 2008: 4)

Menurut Silverus (1991: 80-81) kriteria telaah dari segi materi, konstruksi dan bahasa adalah sebagai berikut:

- 1) Materi
  - a) Kesesuaian soal dengan indikator, apabila soal didasarkan atas kisi-kisi yang memuat indikator soal harus sesuai dengan kisi-kisi.
  - b) Kesesuaian materi yang diukur dengan kompetensi relevansi, kontinuitas, keterpakaian sehari-hari tinggi.
  - c) Pilihan jawaban homogen dan logis.
  - d) Hanya ada satu kunci jawaban.
- 2) Konstruksi
  - a) Pokok soal dirumuskan dengan singkat, jelas, dan tegas.
  - b) Rumusan pokok soal dan pilihan jawaban merupakan pernyataan yang diperlukan saja.
  - c) Pokok soal tidak memberi petunjuk kunci jawaban.
  - d) Pokok soal bebas dari pernyataan yang bersifat negatif ganda.
  - e) Pilihan jawaban homogeny dan logis ditinjau dari segi materi.
  - f) Gambar, grafik, tabel, diagram, atau sejenisnya jelas dan berfungsi.
  - g) Panjang pilihan jawaban relatif sama.
  - h) Pilihan jawaban tidak menggunakan pernyataan “semua jawaban di atas salah/benar” dan sejenisnya.
  - i) Pilihan jawaban yang berbentuk angka waktu disusun berdasarkan urutan besar kecilnya angka atau kronologisnya.
  - j) Butir soal tidak bergantung pada jawaban soal sebelumnya.
- 3) Bahasa
  - a) Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia.
  - b) Menggunakan bahasa yang komunikatif.
  - c) Tidak menggunakan bahasa yang berlaku setempat/tabu.
  - d) Pilihan jawaban tidak mengulang kata/kelompok kata yang sama, kecuali merupakan satu kesatuan pengertian.

**Analisis Butir Soal Secara Kuantitatif**

Ada dua teknik untuk melakukan analisis butir soal secara kualitatif, yaitu: menggunakan teori tes klasik (*classical test theory*) dan teori tes modern (*modern test theory*) ( Aminah, 2012: 5). Dalam penelitian ini dilakukan analisis butir soal secara kuantitatif menggunakan teori tes klasik. Analisis butir soal secara klasik merupakan suatu proses penelaahan butir soal melalui informasi dari jawaban siswa untuk meningkatkan mutu butir soal yang bersangkutan dengan menggunakan terori tes klasik (Depdiknas, 2008: 8).

1) Reliabilitas

merupakan derajat atau tingkat kepercayaan suatu tes. Tujuan utama menghitung besarnya reliabilitas untuk mengetahui tingkat ketepatan dan konsistensi skor tes. Indeks reliabilitas bernilai antara 0-1. Semakin tinggi nilai koefisien tes (mendekati 1) maka makin tinggi pula tingkat ketepatan dan kepercayaan soal tersebut. Suatu tes dikatakan reliabel apabila beberapa kali pengujian menunjukkan hasil yang relatif sama (Sudjana, 2005: 139). Pendekatan yang digunakan untuk menghitung besarnya koefisien reliabilitas

menggunakan pendekatan konsistensi internal (Depdiknas, 2008: 15).

2) Taraf kesukaran soal

Indek kesukaran item menurut Azwar (2002: 134) merupakan rasio antara penjawab item dengan benar dan banyaknya penjawab item. Butir-butir item tes hasil belajar dapat dinyatakan sebagai butir-butir yang baik, apabila butir-butir item tersebut tidak terlalu sukar dan tidak terlalu mudah dengan kata lain derajat kesukaran item itu adalah sedang (Sudijono, 2005: 370).

Butir-butir item tes hasil belajar dapat dinyatakan sebagai butir-butir yang baik, apabila butir-butir item tersebut tidak terlalu sukar dan tidak terlalu mudah dengan kata lain derajat kesukaran item itu adalah sedang (Sudijono, 2005: 370). Sedangkan menurut Sunarya (tth :15) perangkat tes yang baik, memiliki tingkat kesukaran yang seimbang. Seimbang dalam proporsi penyebaran soal mudah, sedang, dan sukar. Proporsinya bisa 20 % mudah, 60 % sedang, dan 20 % sukar, atau komposisi yang lain misalnya (1:2:1). Soal mudah digunakan untuk memberikan motivasi kerja peserta tes, sedangkan soal yang sukar digunakan pada saat seleksi.

Rumus mencari indeks kesukaran

$$P = \frac{n_i}{N} \quad (\text{Azwar, 2002: 134})$$

Di mana:

- P :Indeks kesukaran
- n<sub>i</sub> :Banyaknya siswa yang menjawab item dengan betul
- N : Jumlah peserta tes yang menjawab item

Mengenai bagaimana cara memberikan penafsiran (interpretasi) terhadap angka indek kesukaran item, Sudijono (2005: 372-373) mengutip pendapat Witherington dalam bukunya berjudul *Psychological Education* yang telah ditunjukkan pada Tabel 2.1.

Tabel 2.1 Klasifikasi Indeks Kesukaran dan Interpretasinya

No	Indek Kesukaran (p)	Interpretasi
1	Kurang dari 0,25	Sukar
2	0,25 - 0,75	Sedang
3	0,75 - 1,00	Mudah

3) Daya pembeda

Menurut Sudijono (2005: 385-386), daya pembeda item adalah kemampuan sebuah butir item tes hasil belajar untuk dapat membedakan antara testee yang berkemampuan tinggi dengan testee yang berkemampuan rendah. Sejalan dengan pendapat Azwar (2002: 137) yang menyatakan indeks diskriminasi item adalah kemampuan item dalam membedakan antara siswa yang mempunyai kemampuan tinggi (diwakili oleh mereka yang termasuk kelompok tinggi) dan siswa yang mempunyai kemampuan rendah (diwakili oleh mereka yang termasuk dalam kelompok rendah).

Indeks daya diskriminasi (D) yang berkisar antara 0,00 sampai 1,0. Pada indeks diskriminasi ada tanda negatif (-) menunjukkan jika sesuatu soal “terbalik” yaitu anak pandai disebut bodoh dan anak bodoh disebut pandai. Azwar mengutip dari Ebel (1979) Klasifikasi daya pembeda yang ditunjukkan pada Tabel 2.2.

Tabel 2.2 Klasifikasi Indeks Diskriminasi dan Interpretasinya

No	Indek Diskriminasi (D)	Interpretasi
1	lebih dari 0,40	Bagus sekali
2	0,30 – 0,39	Lumayan bagus, tapi mungkin masih

		perlu peningkatan
3	0,20 – 0,29	Belum memuaskan, perlu diperbaiki
4	kurang dari 0,20	Jelek dan harus dibuang

#### 4) Pola jawaban soal

Suharsimi (2006: 219-220) menjelaskan bahwa yang dimaksud pola jawaban soal adalah distribusi *testee* dalam hal menentukan pilihan jawaban pada soal bentuk pilihan ganda. Pola jawaban soal digunakan mengetahui taraf kesukaran soal, daya pembeda soal, dan baik tidaknya distraktor. Menurut pendapat Fernandes (1984: 29) yang menyatakan bahwa “*Here a rule of thumb is that an alternative should be eliminated or revised unless one of every 50 student (or 2 %) select it.*” (Rokhyati, 2011: 30). Pendapat tersebut menyebutkan bahwa soal tidak perlu dibuang atau direvisi apabila telah dipilih oleh 2 % peserta tes.

Penelitian ini menggunakan pendapat fernandes yang menyebutkan bahwa distraktor telah berfungsi baik jika telah dipilih oleh 2 % peserta tes. Analisis distraktor hanya dapat diterapkan pada bentuk tes pilihan ganda (tes objektif). Menurut pendapat (Rokhyati, 2011: 30-31) suatu butir soal yang baik adalah butir soal yang kunci jawabannya positif dan distraktornya negatif.

#### 5) Pengambilan Keputusan

Pengambilan keputusan terhadap suatu item dapat dikelompokkan menjadi tiga yaitu diterima, direvisi dan ditolak. Menurut Elvin dan Surantoro (2010: 187) pengambilan keputusan penilaian berdasarkan karakteristik tingkat kesulitan, daya beda dan distribusi jawaban (efektifitas distraktor).

Kriteria keputusan untuk penilaian item soal sebagai berikut:

- Item soal diterima, apabila karakteristik item soal memenuhi semua kriteria. Item soal yang terlalu sukar atau terlalu mudah, tetapi memiliki daya beda dan distribusi pengecoh item yang memenuhi kriteria, butir soal tersebut dapat diterima atau dipilih;
- Item soal direvisi, apabila salah satu atau lebih dari ketiga kriteria karakteristik item soal tidak memenuhi kriteria;
- Item soal ditolak (digugurkan), apabila item soal memiliki karakteristik yang tidak memenuhi semua kriteria.

#### g. Ukuran Pemusatan Data

Ukuran pemusatan data (*Central Tendency*) merupakan salah satu aspek yang paling penting untuk menggambarkan karakteristik umum data tersebut. Ukuran tendensi pusat yang baik menurut Djarwanto (2001: 44) adalah yang dapat memenuhi fungsinya, yaitu sebagai ukuran untuk menunjukkan tendensi pusat dari suatu distribusi dan dapat mewakili seluruh nilai pengamatannya.

Menurut Sudaryono, Saefullah dan Raharjo (2012: 51-52), *tendency central* digunakan untuk membandingkan variabel satu dengan variabel lain pada dua atau lebih kelompok data penelitian. Gambaran ringkas tentang variabel penelitian ini dapat memungkinkan peneliti dengan mudah dan cepat dapat membaca dan menggambarkan keadaan suatu variabel penelitian secara menyeluruh.

Tendensi pusat dikatakan baik menurut Djarwanto (2001 :44) apabila memenuhi syarat-syarat antara lain (1) harus dengan tegas dirumuskan pembentukannya, (2) harus didasarkan pada perhitungan seluruh nilai pengamatan, (3) tidak mempunyai sifat matematis yang abstrak, (4) dapat diperoleh dengan perhitungan yang mudah dan cepat, (5) tidak terpengaruh dengan adanya fluktuasi sampling. Ada tiga ukuran pemusatan data yang sering digunakan, yaitu: Mean, Median dan Modus.

Dispersi data (ukuran variasi), menurut Sudaryono, Saefullah dan Raharjo (2012:81) ukuran dispersi atau ukuran variasi atau ukuran penyimpangan adalah ukuran yang menyatakan seberapa jauh penyimpangan nilai-nilai data dari nilai-nilai pusatnya atau ukuran yang menyatakan seberapa banyak nilai-nilai data yang berbeda dengan nilai-nilai pusatnya. Berikut akan dibahas jenis dispersi data yaitu varians dan simpangan baku (standar deviasi).

### 3. Metodologi Penelitian

Penelitian ini merupakan penelitian pengukuran, yaitu: (1) mengidentifikasi parameter soal Ulangan Akhir Semester (UAS) Genap menggunakan program ITEMAN, (2) mengidentifikasi parameter soal menggunakan *tendency central*.

Sampel penelitian ini adalah pola respon yang terdiri dari seluruh data jawaban siswa, dan kunci jawaban soal Ulangan Akhir Semester (UAS) Genap mata pelajaran Fisika Kelas XI SMA di Kabupaten Boyolali Tahun Pelajaran 2012/2013 serta soal yang digunakan. Pola jawaban siswa berjumlah 1036 data jawaban siswa, dihimpun dari sejumlah SMA di Kabupaten Boyolali.

Dari data yang didapatkan, dibuat menjadi dua kelompok data yaitu data 1 dan data 2. Data 1 merupakan data yang mencakup seluruh data yang didapatkan, dan data 2 merupakan data yang mempunyai skor diatas 20 diperoleh dari pemusatan data. Data dikumpulkan dengan teknik dokumentasi.

Teknik Analisis Data dilakukan dengan pendekatan kualitatif dan kuantitatif. Analisis secara kualitatif dilakukan untuk mengetahui kualitas soal dari aspek materi, aspek konstruksi dan aspek bahasa. Analisis kuantitatif dilakukan menggunakan teori tes klasik, dengan program *MicroCat ITEMAN* versi 3,00. ITEMAN merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik.

#### Analisis Kualitatif

Hasil telaah dirangkum dan ditentukan tingkat karakteristik perangkat tes dengan menghitung persentase validitas dalam butir soal. Untuk menghitung persentase validitas dapat menggunakan rumus dibawah ini.

$$\text{Rumus} = \frac{\text{Jumlah aspek yang sesuai}}{\text{Jumlah seluruh aspek telaah}} \times 100 \%$$

#### Analisis Kuantitatif

Respon siswa terhadap perangkat tes dianalisis dengan menggunakan paket program *MicroCat ITEMAN* versi 3.00 dan program SPSS. Hasil pengolahan data mencakup informasi mengenai *tendency central*, besarnya indeks reliabilitas dan analisis butir soal. Analisis butir soal meliputi indeks kesukaran butir, indeks daya beda, dan efektifitas distraktor.

##### a. *Tendency central*

*Tendency central* dapat diketahui dari hasil *print out* program ITEMAN dan dari program SPSS. *Tendency central* meliputi mean, median, modus, varians dan simpangan baku. Pada penelitian ini membandingkan nilai-nilai dari besaran-besaran *tendency central* pada data 1 dan data 2.

##### b. Reliabilitas

Indeks reliabilitas tes dilihat pada nilai *Alpha Cronbach* pada *Output ITEMAN*, semakin besar nilainya menunjukkan semakin kecil tingkat kesalahan pengukuran. Semakin tinggi koefisien reliabilitas (mendekati 1), maka semakin tinggi pula keajegan atau ketepatannya (Depdiknas, 2008: 15). Reliabilitas soal yang relevan dengan penelitian ini adalah sebesar  $\geq 0,70$ .

##### c. Analisis Butir Soal

##### 1) Indeks Kesukaran Butir Soal

Indeks kesukaran butir soal (p) adalah proporsi peserta tes yang menjawab benar butir soal. Indeks kesukaran

butir soal dalam program *ITEMAN* dinyatakan dalam *prop. Correct*. Besarnya tingkat kesukaran yang diterima untuk menyatakan butir soal yang baik adalah antara 0,25 sampai 0,75. Artinya butir dinyatakan tidak baik apabila memiliki tingkat kesukaran tinggi dan apabila butir tersebut memiliki tingkat kesukaran rendah. Butir soal yang telah dianalisis indeks kesukaran butir soalnya, selanjutnya dibuat persentase indeks kesukaran butir soal. Indeks kesukaran dapat dinyatakan baik atau tidak baik sesuai dengan klasifikasi sebagai berikut.

Tabel 3.2 Klasifikasi Indeks Kesukaran

No.	Persentase Indeks Kesukaran	Klasifikasi
1	> 75 %	Sukar
2	25 % sampai dengan 75 %	Sedang
3	< 25 %	Muda

2) Indeks Daya Beda

Daya pembeda butir soal adalah besarnya korelasi antara rata-rata skor peserta yang menjawab benar pada butir soal dengan rata-rata skor total. Menentukan daya beda butir dapat dilihat dari besar nilai korelasi biserial (*Point Biser*). Butir soal yang memiliki indeks daya beda < 0,2 dinyatakan jelek, 0,20 – 0,29 dinyatakan cukup, 0,30 – 0,39 dinyatakan baik dan butir soal yang indeks beda daya ≥ 0,40 dinyatakan baiksekali..

3) Efektivitas Distraktor

*Proportion Endorsing* pada *print out* hasil *ITEMAN* menunjukkan efektivitas distraktor. Distribusi jawaban (distraktor) yang berfungsi dapat ditentukan dengan kriteria sebagai berikut.

- a) Distraktor dikatakan berfungsi dengan baik atau efektif jika direspon oleh peserta tes minimal 2 % pada pengecoh dan *point biser* pada alternatif jawaban negatif kecuali kunci jawaban.
- b) Distraktor belum berfungsi dengan baik atau tidak efektif, apabila respon siswa memilih pengecoh kurang dari 2 % dan angka *point biser* pada alternatif jawaban bernilai positif selain kunci jawaban.

Distraktor dinyatakan gugur apabila tidak ada peserta tes yang memilih distraktor tersebut. Butir soal yang telah dianalisis keefektifan distraktornya.

**Kriteria Kualitas Perangkat Soal**

Kualitas perangkat soal dapat dilihat dari hasil analisis secara kualitatif dan kuantitatif. Analisis kualitatif digunakan untuk melihat aspek materi, aspek bahasa dan aspek konstruksi. Analisis kuantitatif digunakan untuk mengetahui reliabilitas soal dan analisis butir soal yang melingkupi indeks kesukaran, indeks daya beda dan efektivitas distraktor. Pemaknaan terhadap hasil analisis butir soal dilakukan dengan mempertimbangkan indeks kesukaran, indeks daya beda dan efektivitas distraktor.

**4. HASIL PENELITIAN DAN PEMBAHASAN**

**Analisis kualitatif**

Analisis kualitatif dilakukan dengan telaah soal berdasarkan aspek materi, konstruksi dan bahasa. Setelah dilakukan penelaahan berdasarkan aspek materi, konstruksi, dan bahasa. Terdapat butir soal yang telah memenuhi 100 % dari ketiga aspek tersebut dan terdapat butir soal yang kurang dari 100 % tingkat kesesuaiannya dari salah satu atau dua aspek. pengelompokannya dapat dilihat pada Tabel 4.11.

Tabel 4.11 Pengelompokan Butir Soal Berdasarkan Keterpenuhan Ketiga Aspek Telaah

Jumlah terpenuhi ya aspek telaah (100 %)	No soal	Jumlah
Memenuhi 3 aspek telaah	4, 12, 13, 14, 18, 19, 21, 24, 27, 28, 30, 31, 32, 36, 37, 38, 39, 40	18 butir soal (45 %)
Kurang dari 3 aspek telaah	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 15, 16, 17, 20, 22, 23, 25, 26, 29, 33, 34, 35	22 butir soal (55 %)

**Analisis kuantitatif**

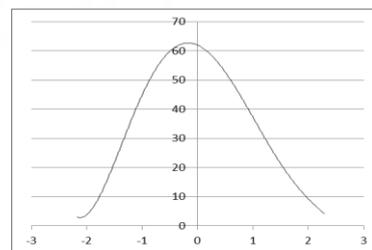
Ukuran *tendency central* meliputi rata-rata hitung (mean), median modus dan dispersi data. Mean, median dan dispersi data dapat diketahui dari *print out* program *ITEMAN* dan SPSS, sedangkan modus dapat diketahui dari jumlah frekuensi yang paling besar atau dapat dilihat pada hasil *print out* SPSS dan pada distribusi skor jawaban siswa pada Gambar 4.1 dan Gambar 4.2.

Perbandingan *tendency central* dan dispersi data pada data 1 dan data 2 dapat dilihat pada Tabel 4.12.

Tabel 4.12 Perbandingan Nilai *Tendency Central* dan Dispersi Data Pada Data 1 dan Data 2

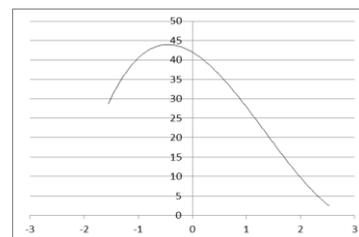
<i>Tendency central</i>	Data 1	Data 2
Mean	21,035	27,23
Median	21,0	27,0
Modus	15,0	28,0
Varians	61,310	21,706
Simpangan Baku	7,830	4,659

4.1 Untuk distribusi data 1 dapat di lihat pada Gambar



Gambar 4.1 Distribusi Data 1

4.2 Untuk distribusi data 2 dapat di lihat pada Gambar



Gambar 4.2 Distribusi Data 2

*commit to user*

Dari Tabel 4.12 diketahui efek dari pemusatan data dari data 1 ke data 2 mengenai nilai *tendency central*. Efek pemusatan tersebut adalah terjadinya pergeseran nilai mean, median dan modus, dimana masing-masing bergeser naik.

Hal ini terjadi karena pemusatan data yang dilakukan ke arah skor yang lebih tinggi. Untuk nilai varians dan simpangan baku (standar deviasi) diketahui nilainya menurun, hal ini merupakan efek pemusatan data. Menurunnya nilai varians dan simpangan baku (standar deviasi) menunjukkan penyimpangan nilai-nilai data terhadap nilai-nilai pusatnya tidak terlalu jauh.

**Analisis Butir Soal**

1) Indeks Kesukaran

Butir item yang baik apabila perbandingan antara jumlah penjawab item dengan benar dan jumlah penjawab item seluruhnya didapat hasil yang sedang atau dalam artian sekitar 0,50. Sehingga item dapat dikatakan tidak terlalu sukar dan tidak terlalu mudah. Besarnya tingkat kesukaran yang diterima untuk menyatakan butir soal yang baik adalah antara 0,25 sampai 0,75.

Data 1 dari hasil *print out* didapatkan besar *mean p* = 0,526 yang menunjukkan bahwa perangkat soal tergolong dalam soal yang baik. Dilihat dari indeks tingkat kesukaran, butir yang termasuk sulit ada 3 butir soal, sedang ada 31 butir soal, dan mudah ada 6 butir soal dengan perbandingan 7,5 % sulit, 77,5 % sedang, dan 15 % mudah. Hasil pengujian tingkat kesukaran tersebut dapat dilihat pada tabel 4.3.

Tabel 4.3 Hasil Pengujian Tingkat Kesukaran Data 1

Tingkat Kesukaran	Butir Soal	Jumlah	Persentase
Sulit (< 0,25)	17, 19, 29	3	7,5 %
Sedang (0,25 – 0,75)	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 14, 15, 16, 18, 20, 21, 22, 23, 25, 27, 28, 30, 31, 33, 34, 35, 36, 37, 38, 40	31	77,5 %
Mudah (> 0,75)	7,12,24,26,32,39	6	15 %

Data 2 dari hasil *print out* didapatkan besar *mean p* = 0,681 yang menunjukkan bahwa perangkat soal tergolong dalam soal yang baik. Dilihat dari indeks tingkat kesukaran, butir yang termasuk sulit ada 0 butir soal, sedang ada 23 butir soal, dan mudah ada 17 butir soal dengan perbandingan 0 % sulit, 57,5 % sedang, dan 42,5 % mudah. Hasil pengujian tingkat kesukaran tersebut dapat dilihat pada Tabel 4.7.

Tabel 4.7 Hasil Pengujian Tingkat Kesukaran Data 2

Tingkat Kesukaran	Butir Soal	Jumlah	Persentase
Sulit (< 0,25)	-	0	0
Sedang (0,25 – 0,75)	1, 4, 5, 6, 8, 9, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 29, 30, 31, 33, 35, 36	23	57,5 %
Mudah (> 0,75)	2, 3, 7, 11, 12, 23, 24, 25, 26, 27, 28, 32, 34, 37, 38, 39, 40	17	42,5 %

2) Indek Daya Beda

Data 1 mempunyai *mean biserial* sebesar 0,542, dengan demikian perangkat soal dapat dikatakan sebagai perangkat soal yang baik sekali. Perangkat soal dapat membedakan peserta tes yang pandai dengan yang kurang pandai (bodoh). Dilihat dari indeks daya beda, butir yang baik sekali ( $\geq 0,40$ ) ada 24 butir soal (60 %). Butir soal yang baik (0,30 – 0,39) ada 11 butir soal (27,5 %). Butir soal yang

cukup (0,20 – 0,29) ada 2 butir soal (5 %). Butir soal yang jelek (< 0,20) ada 3 butir soal (7,5 %). Nilai *point biser* dalam hasil *ITEMAN* tidak ada yang bernilai negatif. Hasil pengujian daya beda dapat dilihat pada Tabel 4.4.

Tabel 4.4 Hasil Pengujian Daya Beda Data 1

Indeks Daya Beda	Butir Soal	Jumlah	Persentase
Jelek (< 0,20)	8, 15, 17	3	7,5 %
Cukup (0,20 – 0,29)	33, 34	2	5 %
Baik (0,30 – 0,39)	1, 4, 5, 7, 12, 21, 22, 29, 30, 37, 39	11	27,5 %
Baik Sekali ( $\geq 0,40$ )	2, 3, 6, 9, 10, 11, 13, 14, 16, 18, 19, 20, 23, 24, 25, 26, 27, 28, 31, 32, 35, 36, 38, 40	24	60 %

Data 2 mempunyai *mean biserial* sebesar 0,374, dengan demikian perangkat soal dapat dikatakan sebagai perangkat soal yang baik. Perangkat soal dapat membedakan peserta tes yang pandai dengan yang kurang pandai (bodoh). Dilihat dari indeks daya beda, butir yang baik sekali ( $\geq 0,40$ ) ada 4 butir soal (10 %). Butir soal yang baik (0,30 – 0,39) ada 14 butir soal (35 %). Butir soal yang cukup (0,20 – 0,29) ada 10 butir soal (25 %). Butir soal yang jelek (< 0,20) ada 12 butir soal (30 %). Nilai *point biser* dalam hasil *ITEMAN* tidak ada yang bernilai negatif. Hasil pengujian daya beda dapat dilihat pada Tabel 4.8.

Tabel 4.8 Hasil Pengujian Daya beda Data 2

Indeks Daya Beda	Butir Soal	Jumlah	Persentase
Jelek (< 0,20)	1, 4, 5, 7, 8, 17, 24, 26, 30, 32, 34, 39	12	30 %
Cukup (0,20 – 0,29)	10, 11, 12, 21, 23, 25, 29, 33, 37, 38	10	25 %
Baik (0,30 – 0,39)	2, 3, 6, 9, 13, 14, 15, 16, 20, 27, 28, 31, 36, 40	14	35 %
Baik Sekali ( $\geq 0,40$ )	18, 19, 22, 35	4	10 %

Dengan melihat hasil dan pembahasan indeks daya beda pada data 1 dan data 2, perangkat soal dapat dikatakan perangkat soal yang baik.

3) Efektivitas Distraktor

Data 1 distraktor yang berfungsi dengan baik ada 138 distraktor (86,25 %). Distraktor yang tidak berfungsi ada 22 distraktor (13,75 %) dan tidak ada distraktor yang digugurkan karena tidak ada distraktor yang tidak dipilih oleh peserta tes.

Data 2 distraktor yang berfungsi dengan baik ada 109 distraktor (68,125 %). Distraktor yang tidak berfungsi ada 46 distraktor (28,75 %) dan distraktor yang digugurkan karena tidak ada distraktor yang tidak dipilih oleh peserta tes ada 5 distraktor (3,125 %).

4) Pemaknaan Hasil Analisis Butir Soal

Dari semua hasil analisis kuantitatif, indeks kesukaran, indek daya beda dan efektivitas distraktor maka perangkat soal dapat dibagi menjadi tiga kriteria yaitu soal yang diterima tanpa ada revisi, soal harus direvisi dan soal digugurkan. Soal digugurkan karena indeks kesukaran termasuk mudah atau sukar, indeks daya beda termasuk jelek dan distraktor harus direvisi.

Menurut data 1 ada 26 (65 %) butir soal yang termasuk soal diterima, 13 (32,5 %) butir soal yang harus direvisi dan 1 (2,5 %) butir soal yang digugurkan. Hasil pengelompokan ini dapat dilihat pada Tabel 4.6.

Tabel 4.6 Pengelompokan Butir Soal Hasil Analisis Data 1

Kondisi soal	Butir Soal	Jumlah	Persentase
Diterima	1, 3, 5, 6, 7, 8, 9, 10, 11, 16, 18, 20, 21, 23, 25, 26, 27, 30, 31, 32, 33, 35, 36, 38, 39, 40	26	65 %
Direvisi	2, 4, 12, 13, 14, 15, 19, 22, 24, 28, 29, 34, 37	13	32,5 %
Digugurkan	17	1	2,5 %

Menurut data 2 ada 10 (25 %) butir soal yang termasuk soal diterima, 24 (60 %) butir soal yang harus direvisi dan 6 (15 %) butir soal yang digugurkan. Hasil pengelompokan ini dapat dilihat pada Tabel 4.10.

Tabel 4.10 Pengelompokan Butir Soal Hasil Analisis Data 2

Kondisi soal	Butir Soal	Jumlah	Persentase
Di terima	3, 9, 10, 16, 20, 21, 22, 36, 40	9	22,5 %
Di revisi	1, 2, 4, 5, 6, 8, 11, 12, 13, 14, 15, 17, 18, 19, 23, 25, 27, 28, 29, 30, 31, 33, 35, 37, 38	25	62,5 %
Digugurkan	7, 24, 26, 32, 34, 39	6	15 %

## 5. KESIMPULAN

(1) karakteristik soal secara kualitatif terdapat 18 butir soal 45 % telah memenuhi ketiga aspek telaah dan 22 butir soal 55 % kurang dari tiga aspek telaah yang terpenuhi, secara kuantitatif reliabilitas soal tergolong baik untuk ilmu alam; indeks kesukaran soal tergolong sedang, Mean p pada hasil ITEMAN data 1 menunjukkan angka 0,526 dan data 2 menunjukkan angka 0,681; daya beda setiap butir soal dinyatakan baik, Nilai point biserial dalam hasil ITEMAN tidak ada yang bernilai negatif. Selain itu hasil print out program ITEMAN menunjukkan angka Mean biserial data 1 (0,542) dan data 2 (0,374); pola jawaban soal berfungsi dengan baik diketahui dari data 1 terdapat 138 distraktor 86,25 % berfungsi efektif dan data 2 terdapat 109 distraktor 68,125 % berfungsi efektif. (2) Efek pemusatan data terhadap parameter besaran pengukuran dengan teori tes klasik dari data 1 ke data 2 yaitu nilai reliabilitas dari 0,879 menjadi 0,678, nilai mean p dari 0,526 menjadi 0,681, nilai mean biserial dari 0,542 menjadi 0,374, nilai mean dari 21,035 menjadi 27,230, median dari 21 menjadi 27, modus dari 15 menjadi 28. varians dari 61, 310 menjadi 21,706, simpangan baku bergeser dari 7,830 menjadi 4,659.

Surakarta, 1 juli 2014  
Mengetahui

Pembimbing 1

Pembimbing 2

Dr. Nonoh Siti Aminah, M.Pd  
NIP. 19510401 197603 2 001

commit to user

Drs. Surantoro, M.Si  
NIP. 19570820 198601 1 001

## 6. DAFTAR PUSTAKA

- Aminah, Nonoh Siti. (2012). *Dasar-dasar Pengukuran dan Statistik Pada Pembelajaran Fisika Cet 1*. Surakarta: UPT UNS Press
- Azwar, Saifuddin. (2002). *Tes Prestasi (Fungsi dan Pengembangan Pengukuran Prestasi Belajar)*. Yogyakarta: Pustaka Pelajar
- Cartono Utari dan Toto Sutarto G. 2006. *Penilaian Hasil Belajar Berbasis Standar*. Bandung : Prisma Press
- Depdiknas. (2008). *Panduan Analisis Butir Soal*. Pembinaan Direktorat SMA
- Djarwanto. (2001). *Mengenal Beberapa Uji Statistik dalam Penelitian*. Yogyakarta : Liberty
- Elvin, Y.E dan Surantoro. (2010). *Pemanfaatan Telaah Kualitatif dan Kuantitatif Sebagai Prasyarat Penyeleksian Item Tes Pilihan Ganda Dalam Penyelesaian Tugas Akhir (Skripsi) di LPTK*. Jurnal AKADEMIKA. Surakarta: UNS
- Haryati, Mimin. (2007). *Model dan Teknik Penilaian pada Tingkat Satuan Pendidikan*. Jakarta: Gaung Persada Press
- Indrastoeti, Jenny SP. (2012). *Pengembangan Asesmen Pembelajaran Sekolah Dasar*. Surakarta : UNS Press
- Purwanto, M. Ngalim. (1988). *Prinsip-prinsip dan Teknik Evaluasi Pengajaran Ed 6 cet 2*. Bandung: Remaja Rosdakarya.
- Rokhyati. (2011). *Karakteristik Secara Kualitatif dan Kuantitatif Soal Ulangan Akhir Semester Genap Bahasa Indonesia Kelas XII SMA Negeri Di Kabupaten Purbalingga Tahun Pelajaran 2010/2011*. Pdf. Diakses pada 28 Januari 2013 Yogyakarta: UNY
- Silverius, Suke. (1991). *Evaluasi Hasil Belajar dan Umpan Balik*. Jakarta: PT Grasindo
- Sudaryono, Asep Saefullah dan Untung Rahardja. (2012). *Statistik Deskriptif For IT Langkah Mudah Analisis Data*. Yogyakarta: Andi
- Sudijono, Anas. (2005). *Pengantar Evaluasi Pendidikan. Ed 1 cet 5*. Jakarta: PT Raja Grafindo Persada
- Sudjana, Nana. (2008). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Suharsimi Arikunto. (2006). *Dasar-dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: PT Bumi Aksara.
- Sunarya, Yaya. (tth). *Pedoman Evaluasi Proses dan Hasil Belajar (Bahan Diskusi di Pusbikte)*. Diakses pada tanggal 28 Januari 2014, dari [http://file.upi.edu/Direktori/FIP/JUR.\\_PSIKOLOGI\\_PEND\\_DAN\\_BIMBINGAN/195911301987031-YAYA\\_SUNARYA/EVALUASI-BAHAN\\_DISKUSI\\_PUSBIKTEK.pdf](http://file.upi.edu/Direktori/FIP/JUR._PSIKOLOGI_PEND_DAN_BIMBINGAN/195911301987031-YAYA_SUNARYA/EVALUASI-BAHAN_DISKUSI_PUSBIKTEK.pdf)
- Tim PEKERTI-AA PPSP LPP Universitas Sebelas Maret. (2007). *Panduan Evaluasi Pembelajaran*. Diakses pada tanggal 20 Desember 2010, dari <http://lpp.uns.ac.id/wp-content/media/PANDUAN-EVALUASI-PEMBELAJARAN.pdf>
- Widoyoko, Eko Putro. (2009). *Evaluasi Program Pembelajaran*. Yogyakarta: Pustaka Pelajar