

**ESTIMASI KESALAHAN BAKU PENGUKURAN SOAL-SOAL UAS FISIKA
KELAS XII SMA DI KABUPATEN BIMA NTB**

Syahrul Ramadhan, Djemari Mardapi
Universitas Negeri Yogyakarta
laarul89@gmail.com

Abstrak

Penelitian ini bertujuan: (1) Mengetahui karakteristik soal-soal UAS mata pelajaran Fisika kelas XII SMA tahun ajaran 2014/2015; (2) Mengestimasi kesalahan pengukuran yang dirancang oleh guru; dan (3) Menentukan metode yang akurat dalam mengestimasi kesalahan baku pengukuran. Analisis data didasarkan pada respon peserta didik terhadap perangkat tes UAS Mata Pelajaran Fisika kelas XII tahun ajaran 2014/2015 di SMAN 1 Sape, SMAN 2 Sape, SMAN 3 Sape, MAN Sape, SMA Muhammadiyah Sape dan SMA PGRI Sape. Analisis butir soal enam perangkat tes tersebut dilakukan dengan Metode Feldt, Metode *Compound Binomial*, Metode Anava, dan Metode Teori Respons Butir. Indeks reliabilitas perangkat tes SMAN 1, SMAN 2 dan SMAN 3 kurang dari 0,7, sedangkan perangkat tes MAN, SMA Muhammadiyah dan SMA PGRI lebih besar dari 0,7. Hasil analisis estimasi kesalahan baku pengukuran berdasarkan Metode *Compound Binomial* memberikan nilai terkecil dibandingkan metode Feldt, metode Analisis Varian dan metode Teori Respon Butir pada tiap perangkat tes.

Kata kunci: *estimasi, kesalahan pengukuran, perangkat tes.*

***THE ESTIMATION THE STANDARD ERROR OF MEASUREMENT
IN PHYSICS END-OF-SEMESTER TESTS OF SENIOR HIGH SCHOOLS
IN KABUPATEN BIMA, NTB***

Syahrul Ramadhan, Djemari Mardapi
Universitas Negeri Yogyakarta
laarul89@gmail.com

Abstract

This study aims to: (1) find out characteristics of Physics End-of-Semester Tests of Senior High Schools in Kabupaten Bima; (2) estimate the standard error of measurement (SEM) of physics subject tests, designed by teachers; and (3) investigate the accuracy of the various methods in estimating SEM. The data analysis was based on students' responses to end-of-semester tests of physics grade XII in the academic year of 2014/2015 in SMAN 1, SMAN 2, SMAN 3, SMA Muhammadiyah, SMA PGRI and MAN. The analysis of the six tests was done using the Feldt Method, ANAVA Method, and Item Response Theory Method. The reability of the test of SMAN 1, SMAN 2 and SMAN 3 is less than 0.7, and the reability of test of MAN, SMA Muhammadiyah and SMA PGRI is more than 0.7. The results of the analysis of the estimated SEM based on the estimated value of the Compound Binomial Method show the smallest SEM than others Method.

Keywords: *estimation, measurement error, test instrument*

Pendahuluan

Pengukuran merupakan suatu langkah yang harus diambil untuk melaksanakan evaluasi. Ketika akan dilakukan sebuah proses evaluasi, maka hendaknya didahului dengan sebuah proses pengukuran yang cermat. Pengukuran pendidikan merupakan kegiatan melakukan kuantifikasi gejala atau objek. Gejala atau objek tersebut bisa berupa motivasi, prestasi, percaya diri atau prestasi yang semuanya dinyatakan dalam bentuk angka (Mardapi, 2012, P. 1). Angka-angka yang didapat dari hasil pengukuran akan membantu peneliti/guru mengetahui tingkat keberhasilan dari proses pembelajaran yang dilakukan. Tentunya informasi yang didapatkan melalui pengukuran merupakan hasil dari proses analisis data dengan teknik tertentu. Selanjutnya barulah guru melakukan penilaian secara intensif untuk melihat keberhasilan dari proses pembelajaran, apakah angka-angka yang didapatkan melalui pengukuran sudah memenuhi standar keberhasilan atau sebaliknya.

Keberhasilan dalam proses pembelajaran dapat dilihat melalui proses penilaian. Dalam usaha menilai prestasi belajar dari para siswa, biasanya guru melakukan pengukuran dalam bentuk ujian-ujian baik tertulis, praktek maupun lisan. Dari hasil ujian-ujian itu guru memberikan skor, yang lazimnya skor tersebut dalam bentuk angka-angka. Namun kemudian timbul pertanyaan, apakah skor yang didapatkan dari hasil ujian tersebut merupakan skor siswa yang sebenarnya?. Wright (2008, P. 130) menyatakan bahwa *True score = observed score ± measurement error*. (skor sebenarnya = skor hasil pengukuran ± kesalahan pengukuran). Dari persamaan tersebut ada dua kemungkinan yang akan terjadi. Pertama, mungkin skor hasil pengukuran lebih kecil dari skor sebenarnya. Kedua, bisa jadi skor hasil pengukuran lebih besar dari skor sebenarnya. Bila salah satu dari kedua kemungkinan tersebut terjadi, berarti pada proses pengukuran telah terjadi kesalahan.

Menghilangkan semua sumber penyebab terjadinya kesalahan pengukuran sangat sulit, tetapi dapat diusahakan agar kesalahan pengukuran dapat diminimalkan, sehingga perolehan skor dapat mencerminkan kemampuan peserta tes yang sebenarnya. Diantara sumber-sumber kesalahan pengukuran itu, nampaknya yang paling mudah dikontrol adalah faktor alat yang digunakan untuk mengukur. Oleh karena itu, dalam usaha untuk memperkecil kesalahan-kesalahan pengukuran adalah dengan memusat-

kan perhatian dalam membuat alat pengukur atau instrument (soal-soal tes) yang baik.

Salah satu faktor yang paling mendasar dalam kesalahan pengukuran adalah alat ukur itu sendiri. Kesalahan dalam pengukuran merupakan hal yang biasa terjadi dalam pembelajaran, namun hal ini sebisa mungkin harus diminimalisir agar kesalahan pengukuran yang terjadi tidak begitu mempengaruhi keberhasilan pembelajaran. Namun, pada kenyataannya masih banyak guru yang belum mampu membuat instrument (soal) dengan baik. Hal tersebut diperkuat dengan pernyataan Miller (2008, P. 93) bahwa "*Most teachers, administrators, and career-guidance personnel recognize that tests (teacher-developed, commercial developed and standardized) are not perfectly valid or reliable*". Pernyataan tersebut bisa diartikan bahwa banyak guru, administrator, dan penyeleksi pekerjaan menyatakan bahwa tidak ada tes yang sepenuhnya valid dan reliabel.

Membuat soal yang valid dan reliabel adalah salah satu kemampuan yang harus dimiliki oleh seorang guru. Akan tetapi, pada kenyataannya masih banyak guru yang belum bisa membuat soal yang valid dan reliabel, oleh karena itu guru seharusnya menyadari akan kelemahan tersebut sehingga bisa memperbaiki kekurangan dan kelemahannya. Soal yang valid dan reliabel merupakan hal penting bagi guru untuk mendapatkan hasil sebenarnya dari keberhasilan proses pembelajaran. Soal/Instrumen yang tidak baik akan mempersulitkan guru dalam mendapatkan informasi tentang keberhasilan pembelajaran yang telah dilakukan dan menyebabkan kesalahan baku pengukuran. guru harus berusaha keras, belajar, dan mengikuti pelatihan-pelatihan dalam rangka meningkatkan kompetensi guru dalam membuat instrumen yang baik agar soal yang dirancangnya memiliki validitas dan reliabilitas yang tinggi dan memiliki kesalahan baku pengukuran yang sekecil mungkin.

Untuk melakukan pengukuran dibutuhkan sebuah instrumen yang baik. Instrumen yang baik adalah instrumen yang memiliki validitas dan reliabilitas tinggi serta memiliki error yang sekecil mungkin dalam menjangkau informasi tentang keberhasilan proses pembelajaran. Instrumen yang baik akan menghasilkan pengukuran yang akurat dan cermat dalam menggali informasi tentang keberhasilan proses pembelajaran (Azwar, 2013, P.2). Senada dengan pendapat tersebut Mardapi (2008, P.67) menyatakan untuk menghasilkan informasi

yang akurat, maka instrumen dalam pengukuran harus handal, sehingga instrument tersebut mampu menghasilkan kesalahan baku pengukuran yang sekecil mungkin. Oleh karena itu, untuk meminimalisir kesalahan dalam pengukuran diperlukan instrument yang baik.

Banyak metode yang berkembang yang bisa digunakan untuk mengestimasi besarnya kesalahan baku pengukuran. Pada prinsipnya metode ini dapat dikategorikan menjadi dua, yaitu yang dikembangkan berdasarkan Teori Tes Klasik dan Teori Respons Butir. Masing-masing model memiliki asumsi dan formula yang berbeda-beda. Pemilihan model tersebut tergantung dari asumsi yang diterapkan pada seperangkat tes yang akan diestimasi kesalahan pengukurannya.

Ulangan Akhir Semester (UAS) merupakan evaluasi pendidikan yang dilakukan setiap 6 bulan sekali. Pelaksanaan evaluasi semacam ini didasarkan pada sistem yang diterapkan yang menganut sistem semester. Bagi guru kelas XII SMA, Ulangan Akhir Semester I diharapkan dapat untuk melihat kesiapan anak dalam menghadapi Ujian Nasional (UN), karena materi kelas XII merupakan 50% dari total materi soal UN di SMA.

Agar hasil UAS dapat menggambarkan hasil yang sesungguhnya, maka perangkat soal yang digunakan harus memenuhi segala persyaratan bagi sebuah alat ukur yang baik dan teruji dalam berbagai aspek. Permasalahan yang muncul sekarang apakah soal-soal UAS benar-benar merupakan alat ukur yang baik yang mampu mencerminkan kemampuan peserta didik. Banyak faktor yang mempengaruhi keadaan tersebut, diantaranya adalah kesalahan pengukuran. Oleh karena itu, setiap hasil pengukuran harus disertai dengan informasi tentang besarnya kesalahan pengukuran.

Meskipun pelaksanaan UAS telah dilakukan beberapa kali, namun hasil yang diperoleh belum memuaskan. Hal ini ditunjukkan dengan perolehan nilai UAS yang relatif rendah, khususnya pada pelajaran Fisika. Rendahnya nilai UAS kemungkinan karena siswa kesulitan mengerjakan soal UAS. Hal ini mengindikasikan tingkat kesukaran soal UAS tidak sesuai dengan kemampuan peserta, daya beda dan pengecoh dalam soal UAS belum mampu berfungsi dengan baik. Besarnya indeks kesukaran dan daya beda butir tes mempengaruhi indeks reliabilitas tes. Tes yang mudah atau sulit cenderung memiliki nilai reliabilitas yang kecil. Kondisi ini menyebabkan terjadinya

kesalahan pengukuran yang cukup tinggi. Berdasarkan hal tersebut perlu diadakan sebuah penelitian untuk melihat besarnya kesalahan baku dalam pengukuran.

Secara umum guru-guru yang berada di Kabupaten Bima tidak melakukan analisis kesalahan baku pengukuran dalam proses pengukuran yang mereka lakukan. Hal tersebut berdasarkan kegiatan prasurevei yang dilakukan oleh peneliti pada beberapa sekolah di kabupaten Bima sejak tanggal 4 sampai 9 agustus 2014. Sebagian besar guru mengaku tidak mengetahui adanya kesalahan baku pengukuran, padahal analisis kesalahan baku pengukuran merupakan hal yang sangat penting karena untuk mengetahui apakah soal yang telah dibuat sudah memenuhi karakteristik baik sehingga hasil pengukuran bisa menggambarkan kemampuan siswa yang sebenarnya. Jadi, dengan adanya penelitian ini diharapkan dapat memberi kontribusi besar bagi perkembangan dan kemajuan pendidikan SMAN di Kabupaten Bima Propinsi NTB.

Berdasarkan uraian sebelumnya, terdapat beberapa permasalahan yang dapat diidentifikasi yang berhubungan dengan pengukuran adalah sebagai berikut: (1) Soal yang digunakan sebagai instrumen belum benar-benar mengukur skor sebenarnya dari siswa. (2) Guru tidak pernah menganalisis besarnya kesalahan baku pengukuran yang terjadi ketika mengevaluasi hasil belajar siswa. (3) Kemampuan guru dalam membuat soal yang valid dan reliabel masih rendah. (4) Kemampuan guru dalam menganalisis hasil belajar siswa dirasakan kurang maksimal. (5) Alat ukur yang digunakan oleh guru dalam hal ini soal tertulis belum bisa mencerminkan kemampuan sebenarnya yang dimiliki oleh siswa.

Tujuan dari penelitian ini adalah (1) Mengetahui karakteristik soal-soal UAS mata pelajaran Fisika kelas XII SMA tahun ajaran 2014/2015. (2) Mengetahui besarnya estimasi kesalahan baku pengukurun soal-soal UAS mata pelajaran Fisika kelas XII SMA tahun ajaran 2014/2015. (3) Mengetahui metode manakah yang paling akurat dalam mengestimasi besarnya kesalahan baku pengukuran soal-soal UAS mata pelajaran Fisika SMA Tahun Ajaran 2014/2015.

Metode Penelitian

Penelitian ini dikategorikan sebagai penelitian *Ex post facto* karena penelitian ini

mendeskripsikan atau menggambarkan peristiwa yang sudah terjadi, berdasarkan data yang telah dikumpulkan dengan tujuan untuk mengetahui besarnya kesalahan baku pengukuran yang terjadi dalam proses pengujian dengan menggunakan Metode Feldt, Metode *Compound Binomial*, Metode Analisis Varian dan Metode Teori Respons Butir.

Tempat penelitian dalam penelitian Estimasi Kesalahan Pengukuran Pada Soal-Soal UAS kelas XII mata pelajaran Fisika SMA jurusan IPA Kecamatan Sape Kabupaten Bima NTB yang dilaksanakan pada bulan Desember 2014 s/d Januari 2015.

Objek dalam penelitian ini adalah seluruh lembar jawaban siswa Ujian Akhir Semester kelas XII mata pelajaran Fisika SMA pada jurusan IPA yang dirancang oleh guru tahun ajaran 2014/2015. Lembar jawaban siswa yang digunakan dalam penelitian ini diperoleh dari 6 SMA yang berada di Kecamatan Sape, Kabupaten Bima Propinsi NTB yang memiliki perangkat tes yang berbeda pada setiap sekolah. Lembar jawaban siswa tersebut dianalisis untuk melihat besarnya kesalahan baku pengukuran yang terjadi perangkat tes yang dibuat oleh masing-masing guru SMAN/ sederajat.

Teknik pengumpulan data dalam penelitian estimasi kesalahan pengukuran pada soal-soal UAS mata pelajaran Fisika SMA Kecamatan Sape Kabupaten Bima NTB adalah dengan menggunakan dokumentasi. Peneliti langsung mendatangi sekolah-sekolah SMAN/ sederajat di Kecamatan Sape yang menjadi objek penelitian dan menemui kepala sekolah untuk meminta izin agar bisa mengambil data di sekolah yang bersangkutan, kemudian peneliti akan menemui guru mata pelajaran Fisika yang mengetahui banyak tentang soal dan lembar jawaban siswa yang telah diujikan pada tahun ajaran 2014/2015. Data yang diambil adalah lembar jawaban siswa dan soal pada Ujian Akhir Semester tahun ajaran 2014/2015.

Analisis kuantitatif karakteristik butir soal pada perangkat tes UAS Fisika SMA dilakukan berdasarkan pendekatan Teori Tes klasik. Analisis tersebut dilakukan untuk melihat informasi tentang butir soal yang layak dan tidak layak untuk diujikan berdasarkan parameter butir, yaitu tingkat kesukaran, daya beda, dan keberfungsian pengecoh.

Analisis tingkat kesukaran perangkat tes UAS Matematika SMAN sederajat dilakukan dengan menggunakan program komputer *MicroCatIteman*. Tingkat kesukaran soal dapat

dilihat pada kolom *Prop. Correct*. Butir soal yang memiliki tingkat kesukaran baik berada dalam interval 0,3 sampai dengan 0,8.

Analisis daya beda perangkat tes UAS Matematika SMAN sederajat dapat dilihat pada kolom *Point Biserial* yang dilakukan menggunakan program komputer *MicroCatIteman*. Kriteria butir soal yang baik memiliki nilai $D \geq 0,3$, sedangkan butir soal yang memiliki nilai $D \leq 0,3$ perlu dilakukan revisi atau diganti dengan butir soal yang baru.

Informasi keberfungsian distraktor juga dapat diperoleh pada program komputer *MicroCatIteman*, yaitu pada kolom *Prop Endorsing*. Distraktor atau pengecoh dikatakan berfungsi apabila nilai dari *Prop Endorsing* pada setiap pilihan ganda memiliki nilai yang lebih besar dari nilai 0,05. Nilai proporsi dari opsi setiap butir soal yang memiliki nilai kurang dari nilai *Prop Endorsing*, maka distraktor tersebut perlu direvisi.

Langkah pertama model ini adalah membuat table distribusi sekor dari hasil tes yang diberikan kepada subjek penelitian satu kali, kemudian dilanjutkan membelah tes secara acak menjadi dua bagian tidak sama panjang dan isinya tetap homogeny (konjenerik). Dengan demikian diperoleh dua distribusi sekor. Masing-masing belahan terdiri dari jumlah butir soal yang berbeda. Untuk mendapatkan indeks reliabilitas tes dengan cara pembelahan tersebut diatas Feldt mengusulkan persamaan sebagai berikut :

$$r_{xx'} = \frac{4(Sx_1x_2)}{s^2x - \left[\frac{(s^2x_1s^2x_2)}{sx} \right]^2}$$

Keterangan :

$R_{xx'}$: Reliabilitas tes

$S_{x_1}^2$: Varian Skor pada belahan 1

$S_{x_2}^2$: Varian Skor pada belahan 2

$S_{x_1x_2}$: Kovarian Skor pada belahan 1 dan 2

S_x : Simpangan baku skor tes

Dengan demikian untuk mencari varian kesalahan terlebih dahulu harus dicari varian skor murni. Untuk mencari varian skor murni dapat digunakan persamaan teori skor murni klasik, $S_e^2 = S_x^2 \cdot R_{xx'}$. Oleh karena itu estimasi varian kesalahan dapat ditentukan sebagai berikut :

$$S_e^2 = S_x^2 - S_t^2$$

Keterangan:

S_e^2 : Estimasi varian kesalahan pengukuran

S_x^2 : Varian skor amatan

S_t^2 : Varian skor murni

Kesalahan pengukuran baku adalah

$$S_e = \sqrt{S_e^2}$$

Metode *Compound Binomial* lebih menekankan pada penyeleksian butir dengan membagi tes menjadi beberapa belahan. Proses penyeleksian butir dilakukan berdasarkan parameter butir tingkat kesukaran soal yang beresetara. Tingkat kesukaran dicari dengan menggunakan bantuan program computer *MicroCat Iteman*. Setelah membagi butir tes menjadi beberapa bagian, langkah selanjutnya adalah mengestimasi kesalahan baku pengukuran berdasarkan Metode *Compound Binomial* dengan menggunakan Rumus.

$$S_{E(i)} = \left[\sum_{h=1}^c \frac{X_{ih}(k_h - X_{ih})}{k_h - 1} \right]^{1/2}$$

Keterangan rumus:

$S_{E(i)}$: Kesalahan pengukuran bagi peserta tes i

X_{ih} : Skor yang diperoleh peserta tes ke i pada rumpun atau kelompok dari butir tes pada kegegori h pada sebuah spesifikasi perangkat tes.

k_h : Jumlah butir soal pada kategori h

Konsep yang digunakan dalam metode ini adalah memandang skor soal dianggap sebagai desain faktorial dua jalan tanpa replikasi. Setiap butir soal dianggap sebagai *treatment* atau perlakuan. Langkah pertama yang dilakukan adalah mencari nilai *mean* kuadrat interaksi peserta tes dan butir soal. Syaifuddin Azwar (2014:93) mengemukakan rumus sederhana dalam menentukan besarnya kesalahan baku pengukuran sebagai berikut:

$$MS_{Sxl} = \frac{\left(i - \frac{\sum X^2}{k} - \frac{\sum Y^2}{n} \right) + \frac{\sum i^2}{nk}}{(n-1)(k-1)}$$

Keterangan rumus:

MS_{Sxl} : *Means Square* interaksi antara peserta tes dan butir soal

X : Nilai yang diperoleh masing-masing peserta tes pada keseluruhan butir soal

Y : Nilai keseluruhan peserta tes pada satu butir soal

Selanjutnya adalah mengestimasi kesalahan baku pengukuran dengan Metode Analisis Varians dengan menggunakan Rumus

$$S_E = [k(MS_{Sxl})]^{1/2}$$

Keterangan rumus:

S_E : Kesalahan Pengukuran

K : Jumlah Butir Soal Sebuah perangkat tes

MS_{Sxl} : *Means Square* interaksi antara peserta tes dan butir soal

Estimasi kesalahan pengukuran berdasarkan Teori Respon Butir dilakukan dengan bantuan Program Komputer *Bilog_MG*. Program tersebut akan menghasilkan nilai paramter butir soal, yaitu tingkat kesukaran, daya beda, dan *guessing*. Langkah utama yang dilakukan mengestimasi kesalahan baku pengukuran berdasarkan Metode Teori Respons Butir adalah menghitung nilai fungsi informasi perangkat tes berdasarkan kemampuan tertentu (θ). Perhitungan nilai fungsi informasi dilakukan berdasarkan pada masing-masing model logistik yang digunakan.

Rumus yang digunakan untuk model logistik satu parameter adalah sebagai berikut.

$$I_i(\theta) = \frac{D^2 e^{D(\theta-b)}}{[1 + D^2 e^{D(\theta-b)}]^2}$$

Perhitungan nilai fungsi informasi bagi perangkat tes dengan model logistik dua parameter adalah sebagai berikut.

$$I_i(\theta) = \frac{D^2 a_j^2 e^{Daj(\theta-bj)}}{[1 + e^{Daj(\theta-bj)}]^2}$$

Perhitungan nilai fungsi informasi bagi perangkat tes dengan model logistik tiga parameter adalah sebagai berikut.

$$I_i(\theta) = \frac{(1-c_j)D^2 a_j^2 e^{2Daj(\theta-bj)}}{[c_j + e^{Daj(\theta-bj)}] [1 + e^{-Daj(\theta-bj)}]^2}$$

Keterangan rumus:

$I(\theta)$: Fungsi informasi suatu perangkat tes.

b_j : Parameter indeks kesulitan butir ke-i

a_j : Parameter daya beda butir ke-i

c_j : Parameter *guessing* daya tebakan butir ke-i

e : Bilangan transenden yang besarnya mendekati 2,718

D : Nilai distribusi logistik besarnya 1.7

Nilai fungsi informasi sangat dibutuhkan dalam mencari kesalahan baku pengukuran dalam Teori Respons Butir (Hambleton, Swaminathan, & Rogers, 1991: 37). Secara sistematis rumus tersebut dapat ditulis :

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

Keterangan rumus:

$SE(\theta)$: Kesalahan baku dalam pengukuran

$I(\theta)$: Harga fungsi informasi tes terhadap paramater tingkat kemampuan peserta tes.

θ : Tingkat kemampuan peserta tes

Hasil Penelitian dan Pembahasan

Analisis Karakteristik

Analisis kuantitatif karakteristik butir soal perangkat tes berdasarkan Teori Tes Klasik dilakukan dengan menggunakan bantuan program komputer *MicroCat Iteman*. Analisis tersebut menghasilkan informasi penting tentang karakteristik butir soal yaitu, reliabilitas, tingkat kesukaran, dan daya beda

Tabel 1. Karakteristik perangkat tes

Sekolah	Karakteristik		
	Reliabilitas	TK	DB
SMAN 1	0,667	0,414-0,800	0,175-0,503
SMAN 2	0,695	0,079-0,911	-0,173-0,659
SMAN 3	0,667	0,077-0,892	-0,277-0,718
MAN	0,818	0,12-0,507	-0,263-0,691
PGRI	0,815	0,193-0,689	0,115-0,684
Muhamadiyah	0,765	0,055-0,80	-0,115-0,729

Analisis Kesalahan Baku Pengukuran

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika SMA dilakukan berdasarkan empat metode estimasi yang berbeda, yaitu Metode Feldt, Metode *Compound Binomial*, Metode Analisis Varians, Metode Teori Respons Butir. Keempat metode tersebut menghasilkan nilai kesalahan baku pengukuran yang bervariasi. Kebervarian nilai estimasi kesalahan baku pengukuran pada metode tersebut disebabkan perbedaan pada langkah-langkah dan landasan rumus yang digunakan dalam menghitung nilai estimasi kesalahan baku pengukuran. Hasil perhitungan estimasi kesalahan baku pengukuran bisa dilihat pada Tabel 2.

Tabel 2. Hasil Analisis SEM

Sekolah	SEM			
	Feldt	Combim	Anava	IRT
SMAN 1	2,507	0,927	2,370	0,441-2,857
SMAN 2	2,115	1,035	1,853	0,479-0,914
SMAN 3	2,621	0,974	2,052	0,396-0,680
MAN	2,694	1,159	2,058	0,422-5,10
PGRI	2,156	0,988	1,997	0,497-2,222
Muhamadiyah	2,504	1,028	2,192	0,447-1,948

Pembahasan

Karakteristik Perangkat Soal

Perangkat tes yang digunakan oleh SMAN 1 terdapat 24 soal yang memiliki kri-

teria tingkat kesukaran yang baik, karena memiliki nilai *prop correct* antara 0,3-0,8. Berdasarkan analisis daya beda, terdapat 19 soal yang memiliki kriteria baik karena memiliki nilai point biser lebih dari 0,3. Berdasarkan analisis keberfungsian distraktor, terdapat 19 soal yang memiliki distractor yang baik. Soal yang memiliki distractor tidak baik disebabkan oleh salah satu atau lebih alternative jawabannya tidak memiliki nilai korelatif negatif dan butir soal alternative jawaban lain yang digunakan lebih tepat dibandingkan kunci jawaban yang telah ditentukan.

Perangkat Tes UAS Fisika SMAN 2 Sape dapat disimpulkan bahwa perangkat tes tersebut terdapat 12 butir soal yang memiliki kriteria tingkat kesukaran yang baik, karena memiliki nilai *prop correct* antara 0,3-0,8. Sedangkan untuk daya beda, terdapat 16 soal yang memiliki kriteria baik karena memiliki nilai point biser lebih dari 0,3. Perangkat selanjutnya yang dianalisis berdasarkan Teori Tes Klasik adalah perangkat soal UAS SMAN 3 Sape. Berdasarkan hasil analisis kuantitatif yang dilakukan dengan menggunakan program *MicroCat Iteman* pada Perangkat Tes UAS Fisika SMAN 3 Sape dapat disimpulkan bahwa perangkat tes tersebut terdapat 22 butir soal yang memiliki kriteria tingkat kesukaran yang baik, karena memiliki nilai *prop correct* antara 0,3-0,8. Sedangkan untuk daya beda, terdapat 20 soal yang memiliki kriteria baik karena memiliki nilai point biser lebih dari 0,3.

Perangkat tes berikutnya adalah perangkat soal UAS MAN Sape. Perangkat tes tersebut terdapat 13 butir soal yang memiliki kriteria tingkat kesukaran yang baik, karena memiliki nilai *prop correct* antara 0,3-0,8. Sedangkan untuk daya beda, terdapat 22 soal yang memiliki kriteria baik karena memiliki nilai point biser lebih dari 0,3. Pada Perangkat Tes UAS Fisika SMA PGRI Sape dapat disimpulkan bahwa perangkat tes tersebut terdapat 16 butir soal yang memiliki kriteria tingkat kesukaran yang baik, karena memiliki nilai *prop correct* antara 0,3-0,8. Sedangkan untuk daya beda, terdapat 16 soal yang memiliki kriteria baik karena memiliki nilai point biser lebih dari 0,3.

Berdasarkan hasil analisis kuantitatif yang dilakukan dengan menggunakan program *MicroCat Iteman* pada Perangkat Tes UAS Fisika SMA Muhammadiyah Sape dapat disimpulkan bahwa perangkat tes tersebut terdapat 22 butir soal yang memiliki kriteria tingkat kesu-

karan yang baik, karena memiliki nilai *prop correct* antara 0,3-0,8. Sedangkan untuk daya beda, terdapat 20 soal yang memiliki kriteria baik karena memiliki nilai point bisser lebih dari 0,3.

Perhitungan dari rumus tingkat kesukaran akan menghasilkan nilai dalam interval 0 sampai dengan 1. Hal tersebut sesuai dengan pendapat Reynolds, Livingston, & Willson (2010, P.149) yang menyatakan bahwa “*the item difficulty index can range from 0,0 to 1,0 with easier item having larger decimal values and difficult item’s at lower value. An item answered correctly by all students receives an item difficulty of 1.0 whereas as item answered incorrectly by oll student receives an item difficulty of 0.0*”. Pengertian tersebut dapat diartikan bahwa indek tingkat kesukaran berada antara interval 0 sampai dengan 1, di mana item lebih mudah mempunyai nilai desimal lebih besar dan item sulit mempunyai nilai desimal lebih kecil. Sebuah item dijawab dengan benar oleh semua siswa menerima sebuah tingkat kesukaran 1, sedangkan semua item dijawab salah oleh peserta didik menerima tingkat kesukaran 0. Berdasarkan pendapat tersebut dapat disimpulkan bahwa ketika nilai desimal dari tingkat kesukaran mendekati 0, maka soal itu sangat sulit, begitu juga sebaliknya jika nilai desimal dari tingkat kesukaran mendekati nilia 1, maka soal itu dapat disimpulkan adalah soal yang sangat mudah.

Besar atau kecilnya daya pembeda tergantung nilai indek diskriminasinya (*discrimination index*). Nilai indek diskriminasi berkisar antara interval -1 dan 1. Mehrens & Lehmann (1991, P.162) menyatakan bahwa “*This value is usually expressed as a decimal and can range from -1.00 to +1.00. If it has a positive value, the item has positive discrimination*”. Pernyataan tersebut bisa diartikan bahwa nilai yang biasa diekspresikan sebagai indeks diskriminasi adalah rentang bilangan decimal dari -1,00 sampai 1,00. Jika nilainya positif, maka butir soal tersebut memiliki diskriminasi yang positif. Lebih Daya beda soal yang bernilai negatif (-1,00) menunjukkan bahwa soal tersebut dijawab benar oleh kelompok dengan kemampuan rendah. Daya beda yang bernilai nol (0) menunjukkan bahwa soal tersebut tidak memiliki daya beda, karena bisa dijawab benar oleh kelompok dengan kemampuan tinggi dan kelompok dengan kemampuan rendah, sedangkan daya beda yang bernilai positif menunjukkan bahwa soal tersebut memiliki

daya beda yang baik, karena hanya bisa dijawab benar oleh siswa-siswa yang memiliki kemampuan tinggi saja.

Kesalahan Baku Pengukuran

Berdasarkan estimasi kesalahan baku pengukuran keempat metode tersebut, rentangan nilai sebenarnya/nilai murni/*true score* untuk masing-masing peserta didik dapat diketahui. Penentuan nilai sebenarnya untuk masing-masing peserta didik dapat ditentukan berdasarkan Metode Teori Tes Klasik. Rumus yang digunakan dalam menentukan nilai *true score* berdasarkan Metode Teori Tes Klasik, yaitu $X=T+E$, X merupakan nilai yang didapatkan melalui proses pengukuran (nilai tampak), sedangkan E adalah kesalahan baku pengukuran yang dihasilkan berdasarkan nilai estimasi kesalahan baku pengukuran yang telah dilakukan keempat metode tersebut, dengan taraf kepercayaan 95%

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika SMAN 1 berdasarkan Metode Feldt sebesar 2,507. Berdasarkan Metode *Compound Binomial* adalah sebesar 0,927. Berdasarkan Metode Analisis Varians adalah sebesar 2,370. Kesalahan baku pengukuran berdasarkan Metode Teori Respons Butir erat kaitannya dengan fungsi informasi, semakin besar fungsi informasi maka semakin kecil kesalahan bakupengukuran. Kesalahan baku pengukuran individu perangkat tes UAS Fisika berdasarkan Model Logistik 1-Parameter dari terkecil hingga terbesar adalah 0,441 s/d 2,857 dengan rata-rata sebesar 0,974.

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika SMAN 2 berdasarkan Metode Feldt sebesar 2,230. Berdasarkan Metode *Compound Binomial* adalah sebesar 1,102. Berdasarkan Metode Analisis Varians adalah sebesar 2,054. Kesalahan baku pengukuran berdasarkan Metode Teori Respons Butir dari terkecil hingga terbesar adalah 0,479 s/d 0,914 dengan rata-rata sebesar 0,663.

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika SMAN 3 berdasarkan Metode Feldt sebesar 3,190. Berdasarkan Metode *Compound Binomial* adalah sebesar 0,960. Berdasarkan Metode Analisis Varians adalah sebesar 2,600. Kesalahan baku pengukuran berdasarkan Metode Teori Respons Butir terkecil hingga terbesar adalah 0,396 s/d 0,680 dengan rata-rata sebesar 0,520.

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika MAN berdasarkan

Metode Feldt sebesar 1,414. Berdasarkan Metode *Compound Binomial* adalah sebesar 1,081. Berdasarkan Metode Analisis Varians adalah sebesar 2,241. Estimasi kesalahan baku pengukuran selanjutnya adalah berdasarkan Metode Teori Respons Butir. Kesalahan baku pengukuran individu perangkat tes UAS Fisika berdasarkan Model Logistik 1-Parameter dari terkecil hingga terbesar adalah 0,422 s/d 5,10 dengan rata-rata sebesar 1,256.

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika SMA PGRI berdasarkan Metode Feldt sebesar 2,058. Berdasarkan Metode *Compound Binomial* adalah sebesar 1,108. Berdasarkan Metode Analisis Varians adalah sebesar 2,200. Kesalahan baku pengukuran berdasarkan Metode Teori Respons Butir dari terkecil hingga terbesar adalah 2,222 s/d 0,497 dengan rata-rata sebesar 0,878.

Estimasi kesalahan baku pengukuran perangkat tes UAS Fisika SMA Muhammadiyah berdasarkan Metode Feldt sebesar 2,662. Berdasarkan Metode *Compound Binomial* adalah sebesar 1,005. Berdasarkan Metode Analisis Varians adalah sebesar 2,246. Kesalahan baku pengukuran berdasarkan Metode Teori Respons Butir dari terkecil hingga terbesar adalah 0,447 s/d 1,948 dengan rata-rata sebesar 0,895.

Berdasarkan hasil analisis Estimasi kesalahan baku pengukuran dengan beberapa metode, terlihat bahwa analisis dengan menggunakan metode *Compound Binomial* menghasilkan estimasi kesalahan baku pengukuran terkecil. Asumsi yang membedakan Metode *Compound Binomial* dibandingkan dengan metode yang lain adalah perakitan perangkat tes bukan berdasarkan pengambilan secara acak butir soal pada sebuah populasi butir soal melainkan dalam bentuk strata. Strata atau tingkatan dalam penyeleksian tersebut dilihat dari tingkat kesukaran suatu soal (*Item Difficulty*). Hasil dari penyeleksian bentuk strata tersebut menyebabkan nilai varian error tiap strata menjadi lebih kecil dibandingkan varian error totalnya, disebabkan skor pada tiap strata hampir identik atau homogen.

Pernyataan tersebut senada dengan yang dikemukakan oleh Feldt, Steffan, dan Gupta (1985, P.354) bahwa "*matching forms during test construction is essentially a process of selecting stratified samples of item rather than completely random samples from the populations of item*". Pengertian tersebut dapat diartikan bahwa pencocokan bentuk konstruksi tes pada dasarnya adalah sebuah proses pe-

milihan sampel item yang bertingkat daripada sampel yang benar-benaracak dari populasi item. Diperkuat dengan pendapat Azwar (2014, P.19) yang menyatakan bahwa semakin besar variabilitas berarti bahwa skor-skor yang ada dalam distribusi tersebut semakin beragam, sebaliknya jika variabilitasnya kecil berarti bahwa skor dalam distribusi itu cenderung sama dan seragam atau disebut homogen. Berdasarkan pernyataan-pernyataan tersebut, bisa disimpulkan bahwa skor yang homogen akan menghasilkan variabilitas yang kecil, sedangkan skor yang heterogen akan menghasilkan variabilitas yang besar.

Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah disajikan dalam BAB sebelumnya, terdapat delapan kesimpulan yang dihasilkan dalam penelitian ini. Kedelapan kesimpulan tersebut adalah sebagai berikut. (1) Indeks reliabilitas perangkat tes SMAN 1, SMAN 2 dan SMAN 3 kurang dari 0,7, sedangkan indeks reliabilitas perangkat tes MAN, SMA Muhammadiyah dan SMA PGRI lebih besar dari 0,7. (2) Berdasarkan Metode Feldt besarnya estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Fisika SMAN 1 Sape, SMAN 2 Sape, SMAN 3 Sape, MAN Sape, SMA PGRI Sape dan SMA Muhammadiyah Sape, secara berturut-turut adalah sebesar, 2,507, 2,115, 2,621, 2,694, 2,156, dan 2,504, kesalahan baku pengukuran terkecil terjadi pada perangkat tes SMAN 2 Sape, sedangkan kesalahan baku pengukuran terbesar terjadi pada perangkat tes MAN Sape. (3) Berdasarkan Metode *Compound Binomial* estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Fisika SMAN 1 Sape, SMAN 2 Sape, SMAN 3 Sape, MAN Sape, SMA PGRI Sape dan SMA Muhammadiyah Sape secara berturut-turut adalah sebesar, 0,927, 1,035, 0,974, 1,159, 0,988, dan 1,028, kesalahan baku pengukuran terkecil berdasarkan Metode *Compound Binomial* adalah perangkat tes SMAN 1 Sape, sedangkan kesalahan baku pengukuran yang terbesar adalah perangkat tes MAN Sape. (4) Berdasarkan Metode Analisis Varians estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Fisika SMAN 1 Sape, SMAN 2 Sape, SMAN 3 Sape, MAN Sape, SMA PGRI Sape dan SMA Muhammadiyah Sape secara berturut-turut adalah sebesar 2,370, 1,853, 2,052, 2,058, 1,997, 2,192, kesalahan baku pengukuran terkecil berdasarkan Metode Analisis Varians

adalah perangkat tes SMAN 2 Sape, sedangkan kesalahan baku pengukuran terbesar adalah perangkat tes SMAN 1 Sape. (5) Berdasarkan Metode IRT estimasi kesalahan baku pengukuran perangkat tes UAS mata pelajaran Fisika SMAN 1 Sape, SMAN 2 Sape, SMAN 3 Sape, MAN Sape, SMA PGRI Sape dan SMA Muhammadiyah Sape secara berturut adalah sebesar 0,569, 0,706, 0,597, 0,673, 0,606, dan 0,583, kesalahan baku pengukuran terkecil berdasarkan Metode IRT adalah perangkat tes SMAN 1 Sape, sedangkan kesalahan baku pengukuran terbesar adalah perangkat tes SMAN 2 Sape. (6) Metode yang paling akurat dalam mengestimasi kesalahan baku pengukuran soal-soal UAS Fisika kelas XII SMA di Kabupaten Bima adalah Metode *Compound Binomial*.

Saran

Berdasarkan hasil penelitian yang telah peneliti lakukan, terdapat beberapa saran yang ingin peneliti ungkapkan sebagai berikut. (1) Dinas Pendidikan di Kabupaten Bima diharapkan mengadakan pelatihan-pelatihan peningkatan kompetensi bagi guru-guru terutama pelatihan pembuatan soal yang baik, yaitu soal yang valid, reliabel, dan memiliki kesalahan baku pengukuran sekecil mungkin. (2) Kepala Sekolah diharapkan mengarahkan guru-guru agar selalu meningkatkan kemampuan pedagogik sebagai pendukung dalam proses pembelajaran dengan mengikuti pelatihan-pelatihan peningkatan kemampuan. (3) Guru diharapkan mengikuti pelatihan-pelatihan peningkatan kompetensi pedagogik guru, agar guru mampu membuat soal yang valid, reliabel, dan memiliki kesalahan baku pengukuran sekecil mungkin, sehingga guru atau pendidik bisa melakukan analisis kesalahan baku pengukuran dari setiap perangkat tes yang telah dirancangnya secara kontiniu. (4) Pengetahuan guru tentang estimasi kesalahan baku pengukuran masih sangat kurang sekali, oleh karena itu penelitian pada responden yang berbeda dalam provinsi atau kabupaten yang berbeda perlu dilakukan.

Daftar Pustaka

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Wadsworth.
- Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes & non tes*. Yogyakarta: Mitra Cendikia
- Djemari Mardapi. (2012). *Pengukuran penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Litera
- Feldt, L. S., Steffen, M., & Gupta, M.C. (1985). A comparison of five method for estimating the standat error measurement at specific score Level. *Applied Psychological Measurement*. Vol. 9. No. 4. Pp. 351-361
- Gardner, J. (2012). *Assessment and learning*. (2rd ed.). Los Angeles: Sage Publications Ltd
- Hambleton, R. K., Swaminathan H., & Rogers H. Jane. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education*. Belmont: Wadsworth Thomson Learning.
- Miller, P. W. (2008). *Measurement and teaching*. Muster: Partric W. Miller & Association.
- Reynolds, C. R., Livingston, R. B., & Wilson, V. (2010). *Measurement and assessment in education*. Upper Saddle River: Pearson.
- Syaifuddin Azwar, (2014). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar
- Wright, R. J. (2008) *Educational assessment*. Thousand Oaks: Sage Publications.