

**KOMPARASI METODE ESTIMASI KESALAHAN PENGUKURAN  
SOAL UJIAN AKHIR SEMESTER MATEMATIKA SMA  
DI KABUPATEN LAMPUNG TENGAH**

Sri Wahyuni, Djemari Mardapi  
Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta  
[sriwahyuni0913@gmail.com](mailto:sriwahyuni0913@gmail.com), [djemarimardapi@gmail.com](mailto:djemarimardapi@gmail.com)

**Abstrak**

Penelitian ini bertujuan: (1) mengetahui besarnya estimasi kesalahan baku pengukuran soal UAS Matematika SMA kelas XII IPA tahun ajaran 2014/2015 di Kabupaten Lampung Tengah, dan (2) mengetahui akurasi metode estimasi kesalahan baku pengukuran dari Metode Thorndike, Metode *Compound Binomial*, Metode Polynomial dan Metode Teori Respon Butir dalam mengestimasi kesalahan baku pengukuran soal UAS Matematika SMA kelas XII IPA tahun ajaran 2014/2015 di Kabupaten Lampung Tengah. Analisis data didasarkan pada respon siswa berupa lembar jawaban dari 876 siswa yang berasal dari 9 SMAN di Kabupaten Lampung Tengah. Metode estimasi kesalahan pengukuran yang digunakan dalam penelitian ini adalah Metode Thorndike, Metode *Compound Binomial*, Metode Polynomial, dan Metode Teori Respon Butir. Hasil analisis data diperoleh bahwa estimasi kesalahan pengukuran dengan menggunakan Metode Thorndike, Metode *Compound Binomial*, Metode Polynomial secara berurutan yaitu 2,6658, 0,9029, 3,9199. Estimasi kesalahan pengukuran dengan Metode Teori Respon Butir diperoleh rentang antara 0,2553 s/d 1,2029. Estimasi kesalahan terkecil yaitu pada  $\theta = 0,9$  dan estimasi kesalahan terbesar pada  $\theta = -3$ . Akurasi metode estimasi kesalahan pengukuran berdasarkan Teori Tes Klasik secara berturut-turut adalah Metode *Compound Binomial*, Metode Thorndike, dan Metode Polynomial.

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

**A COMPARISON OF THE ESTIMATION STANDARD ERROR OF  
MEASUREMENT METHODS IN THE MATHEMATICS FINAL TEST OF  
SENIOR HIGH SCHOOLS IN KABUPATEN LAMPUNG TENGAH**

Sri Wahyuni, Djemari Mardapi  
Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta  
[sriwahyuni0913@gmail.com](mailto:sriwahyuni0913@gmail.com), [djemarimardapi@gmail.com](mailto:djemarimardapi@gmail.com)

**Abstract**

*This study aims to: (1) estimate the standard error measurement of mathematics final test of grade XII of senior high school natural science in the academic year of 2014/2015 in Kabupaten Lampung Tengah; and (2) investigate the accuracy of Thorndike Method, Compound Binomial Method, Polynomial Method, and Item Response Theory Method in estimating standard error measurement of mathematics final test of grade XII of senior high school natural science in the academic year of 2014/2015 in Kabupaten Lampung Tengah. The data analysis was based on students' responses. The data source was computerised answer sheets from 876 students of 9 SMAN in Kabupaten Lampung Tengah. The analysis of estimating SEM the test items was done using the Thorndike Method, Compound Binomial Method, Polynomial Method, and Item Response Theory Method. The results of the analysis based on Methods of Thorndike, Compound Binomial, Polynomial are 2.6658, 0.9029, 3.9199. Based on the Item Response Theory Method, the standard error measurement is found in the interval 0.2553 to 1.2029. The smallest standard error measurement at  $\theta = 0.9$  and the largest standard error measurement at  $\theta = -3$ . The accuracy of the estimated standard error measurement methods based on Classical Test Theory respectively is Compound Binomial Method, Thorndike Method, and Polynomial Method.*

**Keywords:** *estimation, measurement error, test instrument.*

## Pendahuluan

Pendidikan merupakan kunci utama dalam perkembangan sebuah negara. Melalui pendidikan akan tercipta sumber daya manusia yang berkompeten, yang akan bersaing dalam perkembangan ilmu pengetahuan dan teknologi. Saat ini perkembangan ilmu dan teknologi telah berkembang sangat pesat. Sebagai konsekuensinya maka pendidikan memiliki peranan yang sangat penting dalam menghadapi perkembangan ilmu pengetahuan dan teknologi. Perkembangan ilmu pengetahuan dan teknologi pada sebuah Negara akan mencerminkan perkembangan Negara tersebut. Setiap Negara mempunyai hak untuk memajukan pendidikan dalam menghadapi perkembangan ilmu dan teknologi. Pendidikan yang baik akan menghasilkan sumber daya manusia yang cerdas yang mampu menghadapi perkembangan ilmu dan teknologi.

Mencerdaskan kehidupan bangsa adalah salah satu tujuan Negara Indonesia seperti yang tercantum dalam Pembukaan Undang-Undang Dasar 1945. Tujuan tersebut dapat dicapai melalui pendidikan. Pendidikan akan melahirkan manusia yang cerdas sesuai dengan tujuan Negara Indonesia. Namun, dalam pencapaian tujuan tersebut bukan merupakan hal yang mudah. Keberhasilan dari suatu pendidikan harus selalu diukur untuk melihat apakah proses dan hasil pendidikan sudah sesuai dengan tujuan dan Permendiknas yang telah ditetapkan. Hal yang paling dasar yang dilakukan untuk melihat keberhasilan pendidikan adalah dengan melihat keberhasilan proses pembelajaran. Keberhasilan proses pembelajaran dapat dilihat dengan melakukan kegiatan evaluasi, penilaian, dan pengukuran.

Dalam mengetahui keberhasilan proses pembelajaran tidak dapat terlepas dari kegiatan evaluasi, penilaian, dan pengukuran. Keberhasilan proses pembelajaran dapat dilihat melalui hasil evaluasi yang dilakukan oleh guru. Evaluasi proses pembelajaran selalu didahului dengan kegiatan pengukuran. Pengukuran ini kemudian dilanjutkan dengan penilaian dan pada akhirnya akan memperoleh nilai. Nilai inilah yang mencerminkan keberhasilan proses pembelajaran.

Evaluasi, penilaian, dan pengukuran dalam pembelajaran di sekolah sangat erat kaitannya. Hasil evaluasi menggambarkan keberhasilan dari proses pembelajaran sehingga dapat membantu menentukan kebijakan untuk memperbaiki kualitas proses pembelajaran. Pada saat guru atau pendidik memberikan se-

buah tes maka hasil dari tes tersebut merupakan data pengukuran. Data ini merupakan bagian dari penilaian yang selanjutnya akan digunakan untuk mengevaluasi proses pembelajaran. Jadi, hasil evaluasi pembelajaran merupakan hasil kegiatan penilaian dan pengukuran dari pembelajaran.

Pengukuran merupakan salah satu kegiatan yang harus dilakukan dalam evaluasi pembelajaran. Pengukuran sangat berperan penting dalam melakukan evaluasi pembelajaran karena pengukuran dilakukan untuk memperoleh informasi yang menggambarkan karakteristik dari suatu objek yang dievaluasi. Pengukuran akan menghasilkan sebuah angka. Pada dasarnya hasil kegiatan pengukuran yang dinyatakan dalam bentuk angka inilah yang menggambarkan karakteristik atau keberhasilan dari suatu objek yang dievaluasi.

Berdasarkan penjelasan sebelumnya jelas bahwa kegiatan pengukuran merupakan kegiatan yang pertama kali dilakukan dalam melakukan evaluasi. Pengukuran dilakukan dengan menggunakan sebuah instrumen yang telah memiliki karakteristik yang baik agar hasil dari pengukuran benar-benar menggambarkan objek yang diukur. Hasil pengukuran yang cermat akan mencerminkan sebuah informasi yang sebenarnya dari objek yang diukur, misalnya dalam pengukuran hasil belajar siswa, hasil pengukuran hasil belajar siswa mencerminkan informasi yang sebenarnya tentang hasil belajar siswa. Pengukuran harus dilakukan dengan cermat sehingga hasil atau informasi yang diperoleh juga tepat sedangkan pengukuran yang kurang cermat akan memberikan hasil dan informasi yang kurang tepat. Oleh karena itu, hasil pengukuran memberikan kemungkinan bahwa ada perbedaan antara hasil dari pengukuran dengan nilai sebenarnya. Kemungkinan yang muncul adalah hasil pengukuran lebih besar dari nilai sebenarnya atau sebaliknya, hasil pengukuran lebih kecil dari nilai sebenarnya. Kemungkinan tersebut terjadi karena adanya kesalahan pengukuran.

Ditinjau dari sifatnya kesalahan pengukuran dibagi menjadi dua jenis, yaitu kesalahan pengukuran ada yang bersifat acak dan ada yang bersifat sistematis (Mardapi, 2012, p.9). Kesalahan acak disebabkan oleh variasi kondisi fisik dan mental dari peserta didik atau siswa, kondisi lingkungan, materi yang diujikan. Kondisi fisik yang kurang baik, seperti ketika siswa sedang sakit atau siswa merasa cemas (*anxiety*) akan mengakibatkan siswa dalam

mengerjakan soal kurang maksimal. Kondisi lingkungan yang kurang kondusif juga dapat mengakibatkan kesalahan acak, contohnya apabila lingkungan yang digunakan dekat dengan keramaian maka konsentrasi siswa akan terganggu sehingga akan mengakibatkan kesalahan dalam pengukuran.

Kesalahan sistematis disebabkan oleh alat ukur yang digunakan. Kesalahan sistematis terjadi akibat soal atau alat ukur yang dibuat terlalu sulit atau terlalu mudah. Selain itu, kesalahan sistematis juga disebabkan oleh subjektivitas guru atau pengukur, contohnya ketika seorang guru murah dalam memberikan skor maka akan mengakibatkan kesalahan dalam pengukuran. Kesalahan sistematis dapat dikendalikan nilainya, artinya besarnya kesalahan sistematis yang disebabkan oleh alat ukur dapat diestimasi. Sedangkan besarnya kesalahan acak tidak dapat diestimasi, karena untuk mengontrol kondisi fisik dan mental siswa merupakan hal yang sulit untuk dilakukan. Oleh karena itu, kesalahan sistematis lebih mudah diminimalkan dari pada kesalahan acak.

Ada beberapa cara yang dapat dilakukan untuk meminimalkan kesalahan pengukuran. Salah satu cara yang dilakukan untuk meminimalkan kesalahan acak dengan cara membuat suasana yang nyaman dan kondusif, sehingga siswa akan fokus dalam mengerjakan tes. Salah satu cara yang dilakukan untuk mengendalikan kesalahan sistematis yang berasal dari alat ukur yang digunakan adalah dengan membuat instrumen tes yang memenuhi kriteria baik. Instrumen tes yang memenuhi kriteria baik akan memberikan hasil pengukuran yang cermat dan menggambarkan kemampuan yang sebenarnya dimiliki oleh siswa. Oleh karena itu, peranan instrumen tes dalam kegiatan evaluasi hasil belajar siswa sangat penting.

Salah satu peranan instrumen tes adalah untuk mengetahui hasil belajar siswa. Contohnya instrumen tes yang digunakan dalam mengetahui hasil belajar siswa adalah soal Ujian Akhir Semester (UAS). Soal UAS digunakan untuk melihat hasil belajar siswa selama satu semester. Soal UAS yang digunakan harus memiliki kriteria soal yang baik karena hasil UAS akan menggambarkan kemampuan dan hasil belajar siswa selama satu semester. Jika soal UAS yang digunakan telah memenuhi kriteria soal yang baik maka hasil pengukuran juga akan menunjukkan kemampuan siswa yang sebenarnya. Soal UAS yang digunakan oleh beberapa sekolah di beberapa Kabupaten

di Indonesia dibuat oleh Tim MGMP (Musyawarah Guru Mata Pelajaran). Salah satu daerah yang menggunakan soal UAS yang dibuat MGMP adalah Kabupaten Lampung Tengah. Namun, ada juga yang dibuat sendiri oleh guru mata pelajaran di setiap sekolah masing-masing.

Dalam pembuatan soal UAS guru harus memahami dan memiliki kemampuan dalam membuat soal yang memiliki karakteristik baik agar dapat mengurangi kesalahan dalam pengukuran. Karakteristik soal yang baik dapat dilihat melalui validitas, reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian distraktor. Selain itu, guru juga harus dapat menganalisis tentang kesalahan baku pengukuran dari soal yang telah dibuat. Namun, pada kenyataannya guru belum memiliki kemampuan yang baik dalam membuat soal. Hal ini sesuai dengan pernyataan yang disampaikan oleh tim dari Pusat Penilaian Pendidikan Balitbang Kemendikbud Deni Hadiana saat sosialisasi kaidah penyusunan soal ujian bagi para guru di Aula Dinas Pendidikan Provinsi Lampung pada hari Rabu tanggal 9 Oktober 2013. Deni Hadiana menyampaikan bahwa seorang tenaga pendidik idealnya memiliki kecakapan dalam membuat soal sebagai alat evaluasi pembelajaran, namun fakta umum di lapangan menunjukkan hal sebaliknya, kemampuan membuat soal oleh para guru masih di bawah standar. (Lampung Post, 10 Oktober 2013)

Berdasarkan hasil temuan penelitian yang dilakukan oleh Purnomo (2007, p.5) jurusan Sejarah Universitas Negeri Semarang (UNNES) tentang kemampuan guru dalam merancang tes berbentuk pilihan ganda pada mata pelajaran IPS untuk Ujian Akhir Sekolah (UAS) dikemukakan bahwa berdasarkan hasil analisis perangkat tes IPS yang diujikan pada UAS di tiga sekolah dasar di Kecamatan Gajahmungkur Semarang, ternyata sebagian terbesarnya tidak dapat dipakai karena tidak memenuhi persyaratan validitas, reliabilitas, tingkat kesukaran soal dan daya beda soal. Mujimin (2010: p.131) melakukan penelitian di Kecamatan Banyumanik dan memperoleh hasil bahwa soal-soal ujian akhir sekolah mata pelajaran bahasa Jawa tahun 2008/2009 yang disusun oleh perseorangan atau tim di Kecamatan Banyumanik belum memenuhi syarat dilihat dari tingkat kesukaran dan daya beda soal. Hal ini mengindikasikan bahwa guru sekolah dasar belum mampu menyusun soal ujian akhir sekolah untuk mata pelajaran bahasa Jawa.

Fakta-fakta tersebut secara umum menjelaskan bahwa beberapa guru belum memiliki kemampuan yang baik dalam membuat soal. Kemampuan dalam membuat soal yang baik harus dimiliki oleh setiap guru karena melalui soal yang dibuat, guru akan memperoleh informasi tentang pencapaian hasil belajar siswa. Karakteristik soal yang baik dapat dilihat dari parameter tingkat kesukaran, daya beda, dan distraktor. Selain itu, guru juga harus memiliki pemahaman tentang analisis kesalahan baku pengukuran.

Berdasarkan hasil prasarvei di Kabupaten Lampung Tengah diperoleh bahwa soal UAS Matematika khususnya tingkat SMA dibuat oleh Tim MGMP Matematika. Pembuatan soal UAS semester ganjil tahun pelajaran 2014/2015 sudah mulai dibuat di bulan November 2014. Soal yang dibuat berbentuk pilihan ganda oleh guru mata pelajaran Matematika yang ditunjuk sebagai tim, dan dalam Tim MGMP satu butir soal dibuat oleh dua orang guru yang telah ditunjuk. Setelah soal selesai dibuat kemudian seluruh tim menyerahkan soal tersebut ke bagian editor dan dicetak dan siap digunakan untuk UAS. Berdasarkan wawancara yang dilakukan dengan ketua MGMP Matematika SMA menjelaskan bahwa soal UAS yang dibuat oleh MGMP tidak diujicobakan dan tidak dianalisis karakteristik soal seperti validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas distraktor. Selain itu tim MGMP juga tidak pernah melakukan analisis tentang estimasi kesalahan pengukuran. Hal ini disebabkan karena guru belum memahami tentang analisis kesalahan baku pengukuran.

Analisis kesalahan baku pengukuran merupakan hal yang sangat penting karena untuk mengetahui soal yang telah dibuat sudah memenuhi karakteristik baik dan untuk mengetahui apakah hasil dari pengukuran sudah menggambarkan kemampuan siswa yang sebenarnya. Semakin kecil nilai kesalahan pengukuran maka hasil pengukurannya juga menggambarkan kemampuan siswa atau objek yang diukur. Hal ini juga berlaku sebaliknya jika nilai kesalahan pengukuran semakin besar maka akan menimbulkan bias atau penyimpangan dari kemampuan siswa atau objek yang diukur, dengan kata lain hasil pengukuran yang diperoleh tidak menggambarkan kemampuan yang sebenarnya.

Ada beberapa metode yang dapat dilakukan untuk mengestimasi kesalahan baku pengukuran yaitu dengan Teori Tes Klasik dan

Teori Respons Butir, contoh metodenya adalah Metode Thorndike, Metode *Compound Binomial*, Metode Polynomial dan Metode Teori Respons Butir. Metode estimasi kesalahan baku pengukuran berfungsi untuk mengestimasi kemampuan seseorang dengan tingkat kesalahan sekecil mungkin. Asumsi-asumsi yang digunakan dari setiap metode tersebut juga berbeda-beda, sehingga hasil estimasinya juga berbeda-beda.

Berdasarkan uraian di atas pemahaman tentang estimasi kesalahan pengukuran sangat penting bagi guru. Dengan mengestimasi kesalahan pengukuran seorang guru akan mengetahui apakah soal yang dibuat sudah memenuhi kriteria soal yang baik dan dapat mengetahui interval kemampuan sebenarnya yang dimiliki oleh siswa. Dari beberapa metode yang ada, setiap metode menggunakan asumsi-asumsi yang berbeda pula, sehingga hasil estimasinya juga berbeda. Oleh karena itu perlu dilakukan penelitian tentang estimasi kesalahan pengukuran untuk mengetahui metode manakah yang memberikan hasil yang akurat.

beberapa permasalahan yang dapat diidentifikasi yang berhubungan dengan pengukuran adalah kemampuan guru dalam membuat soal masih di bawah standar, soal yang dibuat guru belum memenuhi kriteria soal yang baik, guru belum melakukan analisis tentang karakteristik butir soal, seperti validitas, reliabilitas, tingkat kesukaran, daya beda, dan efektivitas distraktor, guru belum melakukan analisis kesalahan baku pengukuran, dan soal yang dibuat belum mencerminkan kemampuan sebenarnya yang dimiliki siswa.

Tujuan dari penelitian ini adalah mengetahui besar estimasi kesalahan baku pengukuran soal UAS Matematika SMA kelas XII IPA tahun ajaran 2014/2015 di Kabupaten Lampung Tengah, dan untuk mengetahui akurasi metode estimasi kesalahan baku pengukuran dari Metode Thorndike, Metode *Compound Binomial*, Metode Polynomial dan Metode Teori Respons Butir dalam mengestimasi kesalahan baku pengukuran soal UAS Matematika SMA kelas XII IPA tahun ajaran 2014/2015 di Kabupaten Lampung Tengah.

### **Metode Penelitian**

Jenis penelitian ini adalah penelitian deskriptif kuantitatif. Penelitian ini mendeskripsikan besarnya kesalahan baku pengukuran yang dilakukan dengan menggunakan Metode

Thorndike, Metode *Compound Binomial*, Metode *Polynomial* dan Metode *Teori Respons Butir*.

Tempat yang digunakan dalam penelitian ini adalah Kabupaten Lampung Tengah Provinsi Lampung yang dilaksanakan pada bulan Desember 2014 sampai dengan Februari 2015. Objek penelitian ini adalah lembar jawaban siswa dalam UAS Matematika kelas XII pada jurusan IPA, berdasarkan perangkat soal yang dibuat MGMP Matematika tahun ajaran 2014/2015. Lembar jawaban siswa yang digunakan dalam penelitian ini berjumlah 876 yang diperoleh dari 9 SMA yang berbeda di Kabupaten Lampung Tengah provinsi Lampung dengan perangkat tes yang sama pada setiap sekolah.

Teknik pengumpulan data dalam penelitian ini adalah dengan menggunakan teknik dokumentasi. Pelaksanaan penelitian ini dilakukan dengan mendatangi sekolah-sekolah SMAN di Kabupaten Lampung Tengah yang menjadi objek penelitian dan menemui kepala sekolah untuk meminta izin agar dapat mengambil data di sekolah yang bersangkutan, kemudian menemui guru mata pelajaran Matematika yang mengetahui banyak tentang soal dan lembar jawaban siswa yang diujikan pada UAS semester gasal tahun ajaran 2014/2015. Data yang diambil adalah lembar jawaban siswa dan soal pada UAS tahun ajaran 2013/2015 yang dibuat oleh MGMP Matematika di Kabupaten Lampung Tengah.

Teknik analisis data yang dilakukan dalam penelitian ini adalah analisis karakteristik butir soal UAS Matematika SMA dilakukan dengan menggunakan Teori Tes Klasik dan Teori Respon Butir. Kemudian dilanjutkan dengan estimasi kesalahan pengukuran dengan menggunakan Metode Thorndike, Metode *Compound Binomial*, Metode *Polynomial*, dan Metode *Teori Respon Butir*.

Analisis karakteristik butir soal UAS Matematika SMA dilakukan terlebih dahulu dengan menggunakan Teori Tes Klasik. Analisis ini dilakukan untuk mengetahui karakteristik perangkat soal yang dibuat oleh MGMP Matematika SMAN di Kabupaten Lampung Tengah. Analisis ini dilakukan dengan menggunakan program komputer *MicroCat Iteman*. Analisis yang dilakukan yaitu dengan menganalisis tingkat kesukaran, daya beda, dan keberfungsian distraktor atau pengecoh pada pilihan jawaban. Kriteria yang digunakan untuk menentukan tingkat kesukaran didasarkan pada

pendapat Miler (2008, p.131) *a 0.30 to 0,70-difficulty range is acceptable difficulty range*. Reynolds, Livingston, & Willson (2010: p.152) menjelaskan kriteria dari nilai diskriminasi menurut Hopkin dapat dilihat pada Tabel 1.

Tabel 1. Karakteristik Daya Beda

<i>D Value</i>	<i>Kriterion</i>
0,40 and larger	<i>Excellent</i>
0,30 – 0,39	<i>Good</i>
0,11 – 0,29	<i>Fair</i>
0,00 – 0,10	<i>Poor</i>
<i>Negatif Values</i>	<i>Miskeyed or orther major Flaw</i>

Analisis karakteristik butir soal dengan Teori Respons Butir dilakukan dengan menggunakan program *BilogMG*. Analisis dengan *BilogMG* dilakukan untuk melihat kesesuaian perangkat soal dengan tiga model parameter logistik teori respons butir. Dalam penentuan kesesuaian model parameter yang cocok dengan perangkat soal dapat dilihat berdasarkan jumlah butir yang sesuai dengan model tersebut. Untuk melihat kesesuaian atau kecocokan butir dengan model logistik dapat dilihat pada fase dua untuk masing-masing model logistik. Butir soal dikatakan sesuai atau cocok dengan model jika nilai *chi-square* pada masing-masing butir lebih dari 0,05. Selanjutnya, dari hasil analisis *BilogMG* dilihat nilai parameter daya beda (a), tingkat kesukaran (b), dan *pseudo guessing* (c). Nilai untuk masing-masing parameter (a), (b), dan (c) dapat dilihat pada fase dua.

Kriteria yang digunakan untuk melihat butir yang baik dalam Teori Respons Butir digunakan pendapat yang dikemukakan oleh Hadi (2011, p.3) yaitu sebagai berikut:

Tabel 2. Kriteria Butir Soal Berdasarkan Teori Respons Butir

Parameter	Nilai	Keterangan
Daya beda (a)	0,4 s/d 2	a baik
Tingkat kesukaran (b)	-2 s/d 2	b baik
<i>Pseudo guessing</i> (c)	0 s/d 1/k (k=jumlah alternatif jawaban)	c baik
Probabilitas	> 0,05	Fit Model

Langkah yang dilakukan untuk mengestimasi kesalahan pengukuran dengan Metode Thorndike dilakukan dengan formula Teori Tes Klasik yaitu dengan membagi perangkat soal menjadi dua bagian sama panjang. Selanjutnya

pada masing-masing belahan dicari nilai rerata dan nilai varians. Kemudian menghitung nilai varians kesalahan pada masing-masing belahan dengan menggunakan rumus berikut:

$$\sigma_{E1}^2 = \sigma_1 \sqrt{1 - r_{xx}}$$

$$\sigma_{E2}^2 = \sigma_2 \sqrt{1 - r_{xx}}$$

Setelah nilai varians kesalahan pada masing-masing belahan diperoleh selanjutnya mengestimasi kesalahan pengukuran dengan Metode Thorndike dengan menggunakan rumus berikut Feldt, Steffen, dan Gupta (1985, p.352):

$$\sigma_E = \sqrt{\sigma_{E1}^2 + \sigma_{E2}^2}$$

Metode *Compound Binomial* dilakukan dengan membagi soal UAS menjadi beberapa belahan. Pada metode ini pembagian belahan soal didasarkan pada tingkat kesukaran butir soal. Perhitungan tingkat kesukaran soal dilakukan dengan menggunakan program *Microcat IteMan*. Selanjutnya menghitung nilai varians kesalahan pada masing-masing belahan. Langkah selanjutnya adalah mengestimasi kesalahan baku pengukuran dengan menggunakan Metode *Compound Binomial* dengan rumus berikut Feldt, Steffen, dan Gupta (1985, p.354):

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

Keterangan:

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

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

$k_h$  : Jumlah butir soal pada kategori h

Metode Polynomial menggunakan teknik regresi untuk mencari besarnya varians kesalahan. Langkah yang harus dilakukan adalah membelah tes menjadi dua bagian sama panjang. Kuadrat selisih skor dua belahan tes untuk masing-masing individu dianggap sebagai nilai yang diprediksikan oleh teknik regresi dari total skor. Persamaan regresi diperoleh dengan bantuan program SPSS. Setelah diperoleh koefisien persamaan regresi, nilai X (skor tampak) disubstitusikan ke persamaan berikut ini (Feldt, Steffen, dan Gupta, 1985, 353):

$$\hat{Y} = a_0 + a_1 (X) + a_2 (X^2) + \dots + a_k (X^k)$$

Selanjutnya adalah mengestimasi kesalahan baku pengukuran dengan menghitung akar kuadrat  $\hat{Y}$  yang diperoleh dari persamaan tersebut.

Estimasi kesalahan pengukuran dengan menggunakan Teori Respons Butir dilakukan

dengan program komputer *Bilog\_MG*. Dalam program tersebut akan menghasilkan nilai parameter butir soal, yaitu tingkat kesukaran, daya beda, dan *guessing*. Langkah yang dilakukan untuk mencari kesalahan pengukuran terlebih dahulu menentukan model logistik yang cocok dengan perangkat soal UAS Matematika SMA di Kabupaten Lampung Tengah. Selanjutnya, menentukan nilai fungsi informasi pada setiap butir dengan menggunakan nilai  $-3 \leq \theta \leq 3$  dan perubahan  $\theta$  sebesar 0,1. Rumus yang digunakan untuk menghitung fungsi informasi adalah rumus berikut (Hambleton & Swaminathan: 1985, 91):

$$\text{Model 1PL: } I(\theta) = \sum_i D^2 P_i Q_i$$

$$\text{Model 2PL: } I(\theta) = \sum_i D^2 a_i^2 P_i Q_i$$

$$\text{Model 3PL: } I(\theta) = \sum_i D^2 a_i^2 Q_i (P_i - c_i)^2 / (1 - c_i)^2 P_i$$

Keterangan:

$P_i(\theta)$  : probabilitas menjawab benar butir i oleh peserta berkemampuan

$Q_i(\theta)$  : probabilitas menjawab salah butir i oleh peserta berkemampuan  $(1 - P_i(\theta))$

$\theta$  : kemampuan peserta

$e$  : bilangan transenden yang besarnya mendekati 2,718

$a_i$  : daya pembeda butir i

$c_i$  : tebakan (*pseudo guessing*) butir i

$D$  : faktor skala (1,7)

Setelah memperoleh fungsi informasi dari soal kemudian menghitung kesalahan pengukuran berdasarkan Teori Respons Butir dengan menggunakan rumus berikut:

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

## Hasil Penelitian dan Pembahasan

### Hasil Penelitian

Analisis yang dilakukan pertama adalah analisis kuantitatif karakteristik soal berdasarkan Teori Tes Klasik. Analisis ini dilakukan dengan menggunakan bantuan program *IteMan*. Berdasarkan hasil analisis *Microcat IteMan* juga dapat diperoleh informasi reliabilitas perangkat soal yang ditunjukkan oleh nilai *Alpha*. Nilai reliabilitas perangkat soal UAS Matematika SMA di Kabupaten Lampung Tengah adalah 0,805. Mardapi (2012, p.128) besarnya indeks keandalan yang diterima adalah minimal 0,70. Oleh karena itu, dapat disimpulkan bahwa perangkat soal tersebut memiliki reliabilitas yang

baik. Berikut ini adalah rangkuman frekuensi dari masing-masing kriteria tingkat kesukaran, daya beda, dan efektivitas distraktor.

Tabel 3. Distribusi Kategorisasi Parameter Butir Hasil Analisis *IteMan*

Parameter	Kriteria	Frekuensi
Tingkat Kesukaran	Mudah	9 butir
	Sedang	19 butir
	Sulit	2 butir
Daya Beda	Sedang	7 butir
	Baik	8 butir
	sangat baik	15 butir
Efektivitas	Baik	25 butir
Distraktor	kurang baik	5 butir

Selanjutnya, analisis kuantitatif karakteristik berdasarkan Teori Response Butir dilakukan untuk melihat model yang cocok dengan soal UAS Matematika SMA di Kabupaten Lampung Tengah. Analisis kuantitatif karakteristik soal berdasarkan Teori Respon Butir dilakukan dengan menggunakan program komputer *BilogMG*. Berikut hasil analisis model satu parameter logistik (1PL).

Tabel 4. Distribusi Kategorisasi Parameter Butir Model 1PL

Tingkat Kesukaran (b)		
Keterangan	Frekuensi	Persentase (%)
Baik	28	93,33
Tidak Baik	2	6,67
Total	30	100
Kesesuaian Model (Fit Model)		
Keterangan	Frekuensi	Persentase (%)
Fit model	2	6,67
Tidak fit	28	93,33
Total	30	100

Pada analisis model 1PL, parameter yang diketahui adalah parameter tingkat kesukaran (b). Berdasarkan Tabel 3 di atas terdapat dua butir soal yang memiliki tingkat kesukaran tidak baik. Hal ini karena kedua nilai tingkat kesukaran pada kedua butir tersebut tidak berada pada interval  $-2$  sampai dengan  $2$

Setelah analisis model 1PL, langkah selanjutnya adalah analisis model dua parameter logistik (2PL). Pada analisis 2PL parameter yang digunakan adalah tingkat kesukaran dan daya beda. Berikut ini rangkuman distribusi kategorisasi dari hasil analisis model 2PL.

Berdasarkan Tabel 5 diperoleh informasi bahwa butir soal yang memiliki daya beda baik berjumlah 22 sedangkan butir yang me-

iliki daya beda tidak baik berjumlah 8. Kemudian pada parameter tingkat kesukaran terdapat 26 butir yang memiliki kriteria tingkat kesukaran yang baik dan terdapat 4 butir yang memiliki tingkat kesukaran tidak baik. Informasi tentang fit model diperoleh bahwa jumlah butir yang fit model berjumlah 3 butir dan 27 butir tidak fit model.

Tabel 5. Distribusi Kategorisasi Parameter Butir Model 2PL

Daya Beda (a)		
Keterangan	Frekuensi	Persentase (%)
Baik	22	73,33
Tidak Baik	8	26,67
Total	30	100
Tingkat Kesukaran (b)		
Keterangan	Frekuensi	Persentase (%)
Baik	26	13,33
Tidak Baik	4	86,67
Total	30	100
Kesesuaian Model (Fit Model)		
Keterangan	Frekuensi	Persentase (%)
Fit model	3	10
Tidak fit model	27	90
Total	30	100

Langkah selanjutnya adalah analisis model tiga parameter logistik (3PL). Pada model 3PL parameter yang diketahui adalah daya beda (a), tingkat kesukaran (b), dan *guessing* (c) Berikut ini Tabel 6 rangkuman distribusi kategorisasi dari hasil analisis model 3PL.

Tabel 6. Distribusi Kategorisasi Parameter Butir Model 3PL

Daya Beda (a)		
Keterangan	Frekuensi	Persentase (%)
Baik	25	83,33
Tidak Baik	5	16,67
Total	30	100
Tingkat Kesukaran (b)		
Keterangan	Frekuensi	Persentase (%)
Baik	27	90
Tidak Baik	3	10
Total	30	100
Tebakan semu/ <i>Pseudue Guessing</i>		
Keterangan	Frekuensi	Persentase (%)
Baik	20	66,67
Tidak Baik	10	33,33
Total	30	100
Kesesuaian Model (Fit Model)		
Keterangan	Frekuensi	Persentase (%)
Fit model	11	36,67
Tidak fit model	19	63,33
Total	30	100

Berdasarkan Tabel 6 diperoleh informasi bahwa butir soal yang memiliki daya beda baik berjumlah 25 butir sedangkan yang memiliki daya beda tidak baik berjumlah 5 butir. Pada parameter tingkat kesukaran butir yang memiliki kriteria baik berjumlah 27 butir sedangkan yang memiliki kriteria tidak baik berjumlah 3 butir. Selanjutnya pada parameter *c* (*guessing*) diperoleh butir yang memiliki *guessing* baik terdapat 20 butir. Sedangkan, butir yang memiliki *guessing* baik berjumlah 10 butir. Informasi selanjutnya tentang keterangan fit model diperoleh bahwa terdapat 11 butir yang fit model sedangkan butir yang tidak fit model berjumlah 19 butir.

Berdasarkan keseluruhan hasil estimasi kesalahan pengukuran dengan menggunakan metode Thorndike, *Compound Binomial*, Polynomial, dan Teori Respons Butir diperoleh rangkuman yang disajikan pada Tabel 7.

Tabel 7. Hasil Estimasi Kesalahan Pengukuran dari Seluruh Metode

Metode	Estimasi SEM
Thorndike	2,6658
<i>Compound Binomial</i>	0,9029
Polynomial	3,9199
Teori Response Butir	0,2553 s/d 1,2029

Berdasarkan Tabel 7 diketahui bahwa akurasi dari masing-masing metode dalam mengestimasi kesalahan pengukuran diperoleh nilai yang berbeda-beda. Pada metode yang didasari oleh Teori Tes Klasik diperoleh estimasi dari yang terkecil ke yang terbesar secara berurutan yaitu Metode *Compound Binomial*, Metode Thondike, dan Metode Polynomial. Sedangkan metode Teori Respons Butir diperoleh nilai kesalahan pengukuran dalam rentang 0,2553 s/d 1,2029.

#### Pembahasan

Berdasarkan hasil analisis karakteristik yang telah dilakukan diperoleh bahwa pada tingkat kesukaran terdapat 9 butir mudah, 19 butir sedang, dan 2 butir sulit. Pada parameter daya beda terdapat 7 butir memiliki daya beda sedang, 8 butir baik, dan 15 butir sangat baik. Selanjutnya, pada efektivitas distraktor diperoleh bahwa terdapat 25 butir memiliki distraktor yang baik dan 5 butir kurang baik. Berdasarkan keseluruhan parameter tingkat kesu-

karan, daya beda, dan efektivitas distraktor diperoleh 12 butir yang memiliki kriteria baik. Kriteria ini didasarkan pada butir yang memiliki tingkat kesukaran sedang, daya beda baik atau sangat baik, serta efektivitas distraktor baik. Reliabilitas soal yang diperoleh berdasarkan analisis program MicroCat *Iteman* adalah 0,805. Mardapi (2012: p.128) besarnya indeks keandalan yang diterima adalah minimal 0,70. Oleh karena itu, dapat disimpulkan bahwa perangkat soal tersebut memiliki reliabilitas yang baik.

Analisis kuantitatif karakteristik soal berdasarkan Teori Respon Butir dilakukan dengan menggunakan program *BilogMG*. Analisis yang dilakukan yaitu analisis kecocokan model antara perangkat soal UAS Matematika SMA di Kabupaten Lampung Tengah dengan tiga model parameter logistik. Hasil analisis model 1PL diperoleh bahwa terdapat 28 (93,33%) butir memiliki tingkat kesukaran baik dan 2 (6,67%) butir yang memiliki tingkat kesukaran tidak baik. Selanjutnya untuk kriteria fit model terdapat 2 (6,67%) butir soal yang fit model dan 28 (93,33%) butir.

Analisis selanjutnya adalah model 2PL yang melibatkan parameter daya beda dan tingkat kesukaran. Hasil yang diperoleh adalah butir yang memiliki daya beda baik berjumlah 22 (73,33%) butir dan daya beda tidak baik berjumlah 8 (26,67%) butir. Pada parameter tingkat kesukaran terdapat 26 (86,67%) butir yang memiliki tingkat kesukaran baik dan 4 (13,33%) butir yang memiliki tingkat kesukaran tidak baik. kriteria yang digunakan sebagai dasar tingkat kesukaran yang baik sama seperti pada analisis 1PL.

Langkah selanjutnya adalah analisis model 3PL yang melibatkan tiga parameter yaitu daya beda, tingkat kesukaran, dan *guessing*. Pada model 3PL diperoleh bahwa soal yang memiliki daya beda baik berjumlah 25 (83,33%) butir dan soal yang memiliki daya beda tidak baik berjumlah 5 (16,67%) butir. Kemudian diperoleh 27 (90%) butir yang memiliki parameter tingkat kesukaran yang baik dan 3 (10%) butir yang tidak baik. hasil selanjutnya adalah parameter *guessing* yang baik berjumlah 20 (66,67%) butir dan yang tidak baik berjumlah 10 (33,33%) butir. Kesesuaian model dengan soal diperoleh 11 (36,67%) butir yang fit model dan 19 (63,33%) butir tidak fit model.

Berdasarkan hasil dari analisis pada ketiga model logistik diperoleh bahwa perang-

kat soal UAS Matematika SMA di Kabupaten Lampung Tengah memiliki kecocokan model dengan model 3PL. Hal ini didasarkan pada banyaknya jumlah soal yang baik pada hasil analisis model 3PL yaitu 7 butir sedangkan pada model 1PL dan 2PL hanya terdapat 1 butir yang baik. Kecocokan model ini digunakan sebagai dasar dalam melakukan estimasi kesalahan pengukuran dengan Metode Teori Respon Butir.

Pada penelitian ini estimasi kesalahan pengukuran dilakukan dengan empat metode yaitu Metode Thorndike, Metode Compound Binomial, Metode Polynomial, dan Metode Teori Respon Butir. Berdasarkan hasil perhitungan estimasi kesalahan pengukuran pada masing-masing metode yang digunakan nilai yang diperoleh berbeda-beda. Hal ini disebabkan karena asumsi rumus dan langkah-langkah yang digunakan di setiap metode juga berbeda. Metode Thorndike, Metode Compound Binomial, dan Metode Polynomial, merupakan metode estimasi kesalahan pengukuran berdasarkan Teori Tes Klasik, sedangkan Metode Teori Respon Butir merupakan metode estimasi kesalahan pengukuran berdasarkan Teori Tes Modern.

Berdasarkan hasil perhitungan diperoleh bahwa metode yang menghasilkan nilai estimasi kesalahan pengukuran berdasarkan Teori Tes Klasik paling kecil adalah Metode *Compound Binomial* dengan nilai 0,90289. Sedangkan metode yang menghasilkan estimasi kesalahan pengukuran yang terbesar adalah Metode Polynomial yaitu 3,9199. Urutan nilai estimasi kesalahan pengukuran berdasarkan Teori Tes Klasik dari yang terkecil ke terbesar secara berurutan adalah Metode Metode Compound Binomial, Metode Thorndike, dan Metode Polynomial.

Nilai kesalahan pengukuran dapat digunakan untuk mengestimasi interval skor sebenarnya (*true score*) pada masing peserta didik. Hal ini didasarkan pada Teori Tes Klasik bahwa  $X = T + E$ , dengan  $X$  adalah skor tampak atau skor yang diperoleh dari hasil tes,  $T$  adalah *true score* skor sebenarnya, dan  $E$  adalah nilai kesalahan pengukuran (dalam penelitian ini adalah nilai kesalahan pengukuran yang diestimasi oleh Metode Thorndike, Metode Compound Binomial, Metode Polynomial, dan Metode Teori Respon Butir). Penentuan interval *true score* pada penelitian ini menggunakan taraf kepercayaan sebesar 95%.

Estimasi kesalahan pengukuran yang dilakukan pertama adalah estimasi kesalahan pengukuran dengan Metode Thorndike. Pada Metode Thorndike diperoleh nilai kesalahan pengukuran pada perangkat soal UAS Matematika SMA di Kabupaten Lampung Tengah sebesar 2,6658. Nilai tersebut digunakan untuk menentukan interval *true score* peserta tes dengan mensubstitusikan ke dalam persamaan umum Teori Tes Klasik. Interval *true score* peserta tes yang diperoleh dengan nilai kesalahan pengukuran berdasarkan metode Thorndike dan taraf kepercayaan 95% adalah  $X - 5,2250 \leq T \leq X + 5,2250$ .

Metode estimasi kesalahan pengukuran yang selanjutnya dilakukan adalah estimasi kesalahan pengukuran dengan Metode Compound Binomial. Nilai kesalahan pengukuran yang diperoleh dengan menggunakan Metode Compound Binomial adalah 0,9029. Berdasarkan nilai tersebut dengan taraf kepercayaan 95% diperoleh interval *true score* peserta tes yaitu  $X - 1,7697 \leq T \leq X + 1,7697$ .

Langkah selanjutnya adalah estimasi kesalahan pengukuran dengan Metode Polynomial. Nilai kesalahan pengukuran yang diperoleh dengan menggunakan Metode Polynomial adalah 3,9199. Berdasarkan nilai tersebut dengan taraf kepercayaan 95% diperoleh interval *true score* peserta tes yaitu  $X - 7,6830 \leq T \leq X + 7,6830$ .

Estimasi kesalahan pengukuran yang terakhir dilakukan adalah estimasi kesalahan pengukuran dengan Metode Teori Respon Butir. Nilai kesalahan pengukuran yang diperoleh dengan menggunakan Metode Teori Respon Butir pada perangkat soal UAS Matematika SMA di Kabupaten Lampung Tengah dengan model 3PL diperoleh rentang antara 0,2553 s/d 1,2029. Estimasi kesalahan terkecil 0,2553 yaitu pada  $\theta = 0,9$  dan estimasi kesalahan terbesar 1,2029 pada  $\theta = -3$ .

Berdasarkan keseluruhan hasil estimasi kesalahan pengukuran dengan Teori Tes Klasik diperoleh bahwa metode yang menghasilkan akurasi yang berbeda-beda. Nilai estimasi kesalahan pengukuran terkecil adalah Metode Compound Binomial. Hal ini karena pada Metode Compound Binomial perangkat tes dibelah menjadi strata berdasarkan tingkat kesukaran. Feldt, Steffen, dan Gupta (1985, p.354) *matching forms during test construction is essentially a process of selecting stratified samples of items rather than completely random samples from the population of item*. Pembelahan berda-

sarkan strata akan menghasilkan skor pada masing-masing belahan seragam. Hal ini akan menyebabkan variabilitas kecil. Ketika variabilitas skor kecil maka nilai varians error juga kecil sehingga kesalahan pengukuran juga kecil.

Azwar (2014, p.19) semakin besar variabilitas berarti bahwa skor-skor yang ada dalam distribusi tersebut semakin beragam, sebaliknya bila variabilitas kecil berarti bahwa skor-skor dalam distribusi itu cenderung sama dan seragam. Berdasarkan pendapat tersebut maka pada masing-masing belahan strata akan diperoleh nilai varians error yang kecil sehingga estimasi kesalahan pengukuran yang diperoleh juga relatif kecil. Oleh karena itu, estimasi kesalahan pengukuran dengan menggunakan Metode Compound Binomial menghasilkan nilai yang lebih kecil dibandingkan dengan Metode Thorndike dan Metode Polynomial.

## Simpulan dan Saran

### Simpulan

Berdasarkan hasil penelitian dan pembahasan, diperoleh simpulan sebagai berikut: (1) Hasil estimasi kesalahan baku pengukuran pada soal-soal UAS kelas XII mata pelajaran Matematika SMA jurusan IPA Kabupaten Lampung Tengah Provinsi Lampung berdasarkan Metode Thorndike adalah 2,6658; (2) Hasil estimasi kesalahan baku pengukuran pada soal-soal UAS kelas XII mata pelajaran Matematika SMA jurusan IPA Kabupaten Lampung Tengah Provinsi Lampung berdasarkan Metode Compound Binomial adalah 0,9029; (3) Hasil estimasi kesalahan baku pengukuran pada soal-soal UAS kelas XII mata pelajaran Matematika SMA jurusan IPA Kabupaten Lampung Tengah Provinsi Lampung berdasarkan Metode Polynomial adalah 3,9199; (4) Hasil estimasi kesalahan baku pengukuran pada soal-soal UAS kelas XII mata pelajaran Matematika SMA jurusan IPA Kabupaten Lampung Tengah Provinsi Lampung berdasarkan Metode Teori Respon Butir diperoleh rentang antara 0,2553 s/d 1,2029. Estimasi kesalahan terkecil yaitu pada  $\theta = 0,9$  dan estimasi kesalahan terbesar pada  $\theta = -3$ ; (5) Metode yang paling akurat dalam mengestimasi kesalahan pengukuran pada soal ujian akhir semester matematika SMA di Kabupaten Lampung Tengah adalah Metode *Compound Binomial*.

### Saran

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, peneliti ingin memberikan saran sebagai berikut: (1) Soal Ujian Akhir Semester Ganjil Matematika SMA jurusan IPA tahun ajaran 2014/2015 di Kabupaten Lampung Tengah yang dibuat oleh Tim MGMP masih tergolong kurang baik, oleh karena itu Tim MGMP sebaiknya mengoreksi kembali soal sebelum digunakan. Selain itu jika perangkat soal sudah digunakan, sebaiknya Tim MGMP bersama guru dapat menganalisis secara kuantitatif perangkat soal yang telah dibuat sehingga pada akhirnya nanti dapat membuat soal yang baik, dan (2) Tim MGMP Matematika Kabupaten Lampung Tengah dapat menggunakan Metode Thorndike, *Compound Binomial*, Polynomial, dan Teori Respon Butir untuk mengestimasi kesalahan pengukuran dari perangkat soal yang dibuat.

### Daftar Pustaka

- Purnomo, Arif. (2007). Kemampuan guru dalam merancang tes berbentuk pilihan ganda pada mata pelajaran ips untuk ujian akhir sekolah (UAS). Lembaran Ilmu Kependidikan. Jilid 36. No. 1.
- Mardapi, Djemari. (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.
- Kemampuan Guru Membuat Soal Rendah. (10 Oktober 2013). *Lampung Post*, hlm.19.
- Miller, P. W. (2008). *Measurement and teaching*. Muster: Partric W. Miller & Association.
- Mujimin. (2010). Kompetensi guru dalam menyusun butir soal pada mata pelajaran bahasa jawa di sekolah dasar. Lembaran Ilmu Kependidikan Edisi September 2010. 125-132
- Reynolds, C. R., Livingston, R. B., & Wilson, V. (2010). *Measurement and assessment in education*. Upper Saddle River: Pearson.

Azwar, Saifuddin. (2014). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.

Hadi, Samsul. (2014). *Lembar Kerja Teknik Analisis Data-Pascasarjana, Mengolah*

*hasil analisis Bilog dengan SPSS*. Materi Kuliah Praktik Pengukuran dan Pengujian, Pascasarjana UNY.