

ESTIMASI KESALAHAN PENGUKURAN PERANGKAT SOAL UJI COBA UJIAN NASIONAL MATA PELAJARAN EKONOMI SMA DI KABUPATEN BANJARNEGARA

Khotimah Marjiastuti, Nahiyah Jaidi Faraz
Prodi Penelitian dan Evaluasi Pendidikan PPs UNY, Universitas Negeri Yogyakarta
khotimahmarjiastuti@gmail.com, nahiya_jf@yahoo.com

Abstrak

Penelitian ini bertujuan untuk: (1) mengetahui karakteristik perangkat soal uji coba UN mata pelajaran ekonomi SMA; (2) mengestimasi kesalahan pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA dan (3) mengetahui metode yang paling akurat dalam mengestimasi kesalahan baku pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara Tahun Ajaran 2014/2015. Data diperoleh dari respon peserta didik terhadap soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara. Lembar jawab yang dianalisis sebanyak 809 lembar yang diperoleh dari dua paket soal yakni paket 1 dan paket 2. Teknik pengambilan data yang digunakan adalah dokumentasi. Metode yang digunakan untuk mengestimasi kesalahan pengukuran perangkat soal tersebut adalah metode Thorndike, Binomial Lord Modifikasi Keats, Analisis Varians dan Teori Respon Butir. Hasil penelitian menunjukkan bahwa (1) soal paket 1 memiliki 13 butir soal dengan kategori baik dari total soal 40 butir, sedangkan soal paket 2 memiliki 20 butir dengan kategori baik dari total soal 40 butir; (2) estimasi kesalahan baku pengukuran berdasarkan metode Thorndike, Binomial Lord modifikasi Keats, Analisis Varians dan Teori Respon Butir untuk paket satu secara berturut-turut 1,8005; 0,8742; 2,7873 dan 0,1990 s.d 1,0320. Estimasi kesalahan baku pengukuran untuk paket dua secara berturut-turut 1,7673; 0,8392; 2,8302 dan 0,2005 s.d 1,0073; (3) metode yang paling akurat dalam mengestimasi kedua paket soal yaitu metode Binomial Lord Modifikasi Keats.

Kata Kunci: *estimasi kesalahan pengukuran, perangkat tes, uji coba Ujian Nasional.*

THE STANDARD ERROR ESTIMATE OF MEASUREMENT IN NATIONAL EXAMINATION TRIALS ON ECONOMICS AT SENIOR HIGH SCHOOLS IN KABUPATEN BANJARNEGARA

Khotimah Marjiastuti, Nahiyah Jaidi Faraz
Prodi Penelitian dan Evaluasi Pendidikan PPs UNY, Universitas Negeri Yogyakarta
khotimahmarjiastuti@gmail.com, nahiya_jf@yahoo.com

Abstract

This study aims to : (1) know the characteristic of national examination trials on economics of Senior High Schools; (2) estimate the standard error of measurement in national examination trials on economics of Senior High Schools and; (3) know the most accurate method to estimate the standard error of measurement at national examination trials on economics of Senior High Schools in Banjarnegara regency in the academic year 2014/2015. The data obtained from the response of students to the national examination trials of economics senior high schools in Kabupaten Banjarnegara. Total answer sheets were analyzed as many as 809 sheets, divided into two packs of tests which are obtained by documentation technique. The methods used to estimate the measurement error of the test is Thorndike method, Binomial Lord modification Keats method, Analysis of Variance method and Item Response Theory method. The results show that (1) the first package has 13 good items from total 40 items, while the second package has 20 good items from total 40 items; (2) the estimated standard error of measurement based on method Thorndike, Binomial Lord modification Keats, Analysis of Variance and Item Response Theory for the first package is 1.8005; 0.8742; 2.7873 and 0.1990 to 1.0320 respectively. Estimates of standard error of measurement for the second package is respectively 1.7673; 0.8392; 2.8302 and 0.2005 to 1.0073; (3) The most accurate method for both packages that is Binomial Lord modification of Keats method.

Keywords: *estimate the standard error of measurement, test instrument, national examination trials.*

Pendahuluan

Berdasarkan Undang-Undang No.20 Tahun 2003, pendidikan merupakan usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang dimiliki dirinya, masyarakat, bangsa, dan negara. Setiap Warga Negara Indonesia berhak memperoleh pendidikan yang layak, karena pendidikan mendorong kemajuan negara. Pendidikan di Indonesia terkait erat dengan kurikulum, pembelajaran dan evaluasi pendidikan.

Kurikulum merupakan seperangkat rencana dan pengaturan mengenai tujuan, isi dan bahan pelajaran serta cara yang digunakan sebagai pedoman penyelenggaraan kegiatan pembelajaran untuk mencapai tujuan pendidikan tertentu. Pembelajaran adalah proses interaksi peserta didik dengan pendidik dan sumber belajar pada suatu lingkungan belajar. Kegiatan yang tidak kalah penting yaitu evaluasi pendidikan yang dilaksanakan sebagai upaya pertanggungjawaban penyelenggaraan pendidikan. Evaluasi dilakukan oleh pendidik untuk memantau proses, kemajuan, dan perbaikan hasil belajar peserta didik secara berkesinambungan. Kegiatan evaluasi dalam pendidikan tidak akan terlaksana tanpa adanya kegiatan pengukuran dan penilaian. Pengukuran dilaksanakan dengan tujuan memberikan atribut berupa angka pada individu. Dalam dunia pendidikan, kegiatan pengukuran dilaksanakan dengan menggunakan instrumen tes dan nontes. Penggunaan instrumen tes atau nontes disesuaikan dengan ranah yang hendak diukur. Kegiatan selanjutnya yaitu penilaian, yang merupakan pemberian label terhadap seseorang. Dalam dunia pendidikan pemberian label dilakukan dengan memberikan kriteria atas nilai yang diperoleh. Nilai yang diperoleh dapat dikategorikan menjadi baik, cukup baik atau bahkan kurang baik.

Mardapi (2008, p.5) menyatakan, untuk meningkatkan kualitas pendidikan dapat ditempuh dengan dua cara, yaitu meningkatkan kualitas pembelajaran dan kualitas sistem penilaiannya. Kegiatan pengukuran yang cermat akan dapat memberikan informasi yang tepat untuk bahan evaluasi. Akan tetapi, dalam kegiatan pengukuran tidak terpisahkan dari kesalahan baku pengukuran (*Standard Error of*

Measurement/SEM). Kesalahan pengukuran yang dimaksud yaitu nilai/skor hasil pengukuran lebih rendah daripada nilai/skor yang sebenarnya atau bahkan nilai/skor hasil pengukuran lebih tinggi dari nilai/skor yang sebenarnya. Tighe (2010, p.3) menyatakan “*reliability also shows problems when numbers of candidates in examinations are low and sampling error affects the range of candidate ability. SEM is not subject to such problems; it is therefore a better measure of the quality of an assessment and is recommended for routine use*”. Bahwa sering orang mengandalkan kualitas soal hanya pada reliabilitasnya, namun di sisi lain bila sampai salah dalam menentukan sampel dan jumlahnya maka akan mempengaruhi hasil estimasi kemampuan peserta tes.

Analisis SEM dapat menjadi solusi pengukuran yang lebih baik dan disarankan untuk rutin dilakukan. Dengan menganalisis tingkat SEM, maka dapat diketahui kemungkinan skor murni (*true score*) peserta tes, serta mengetahui tingkat kemampuan peserta tersebut berada pada kemampuan tinggi, sedang, atau rendah. Raju, et all (2006:p.1) menyatakan “*the SEMs is also referred to as the conditional (or examinee-level) SEMs because it could vary from one examinee to the next. These conditional SEMs are very helpful in practice in interpreting the closeness between an observed score (xs) and the underlying, unobserved true score (ts) or in establishing confidence intervals*”. Jadi, kualitas perangkat soal tidak hanya tergantung pada reliabilitas soalnya saja. Analisis SEM juga berperan untuk mengestimasi skor murni berdasarkan skor tampak yang diperoleh siswa.

Kesalahan pengukuran dalam bidang pendidikan tidak dapat dihindari karena subyek dan obyek yang terlibat yaitu manusia. Ada dua macam kesalahan pengukuran, yaitu kesalahan acak dan kesalahan sistematis (Mardapi, 2008:p.3). Kesalahan acak disebabkan oleh kondisi fisik dan mental yang diukur dan yang mengukur bervariasi. Kesalahan sistematis terjadi karena alat ukur yang digunakan. Ada pendidik yang memberikan soal terlalu mudah sehingga siswa mendapat skor yang tinggi. Sebaliknya, bila pendidik memberi soal yang sulit, maka siswa akan mendapat skor yang rendah. Apabila hal ini sampai terjadi, maka dapat merencanakan evaluasi yang dihasilkan. Oleh karena itu, alat ukur (tes) memiliki peran yang sangat penting dalam dunia pendidikan sehingga soal tes perlu dipersiapkan sebaik

mungkin dan seminimal mungkin kesalahan yang dihasilkan agar lebih representatif mencerminkan kemampuan peserta didik.

Penilaian yang dilakukan oleh pendidik terdiri atas berbagai jenis, seperti kuis, pertanyaan lisan, tugas individu, tugas kelompok, Ulangan Harian (UH), Ulangan Tengah Semester (UTS), dan Ulangan Akhir Semester (UAS). Semua itu dilakukan agar guru dapat mengetahui kemampuan dan kemajuan peserta didik, bahkan bila perlu dilakukan perbaikan atau pengayaan maka hal itu pun akan dilakukan. Hal tersebut dikarenakan pendidik memiliki tanggung jawab atas peserta yang dididik agar menjadi orang yang pandai, berakhlak mulia, tangkas dan terampil. Penilaian terhadap peserta didik tidak hanya dilakukan oleh guru, tetapi juga oleh pihak independen yaitu Badan Standar Nasional Pendidikan (BSNP) yang bekerja sama dengan instansi terkait di lingkungan Pemerintah, Pemerintah Provinsi, Pemerintah Kabupaten/Kota, dan satuan pendidikan. Penilaian tersebut dikenal dengan nama Ujian Nasional (UN).

Penyelenggaraan UN tahun ini sepenuhnya untuk menilai pencapaian standar kompetensi lulusan pada mata pelajaran tertentu secara nasional, kemudian hasilnya digunakan untuk pemetaan mutu, dasar seleksi masuk jenjang pendidikan selanjutnya, dan untuk pembinaan. Melihat adanya tujuan yang besar dari penyelenggaraan UN walaupun sudah tidak lagi menjadi penentu kelulusan yang utama bagi siswa, tetapi sekolah tetap melaksanakan uji coba UN seperti tahun-tahun sebelumnya. Uji Coba UN bertujuan untuk mempersiapkan siswa menghadapi UN sehingga diharapkan uji coba UN dapat memberikan gambaran terkait materi, jenis, bentuk serta cara pengerjaan UN kelak. Intensitas serta waktu pelaksanaan uji coba UN masing-masing sekolah berbeda-beda tergantung pada kebijakan Kepala Sekolah yang bersangkutan.

Berdasarkan hasil observasi beberapa SMA negeri dan swasta di Kabupaten Banjarnegara, umumnya sekolah mengadakan dua kali uji coba UN dengan soal buatan guru sekolah yang bersangkutan dan satu kali uji coba UN dengan soal buatan MGMP yang diterapkan di seluruh SMA baik sekolah negeri maupun swasta di Kabupaten Banjarnegara. Kenyataan menunjukkan bahwa untuk pelaksanaan uji coba UN, guru masih menggunakan soal-soal uji coba tahun lalu atau mengambil dari buku latihan untuk dijadikan soal *try out* di tahun

berikutnya. Permasalahannya pada soal-soal tersebut belum pernah dilakukan analisis butir soal untuk menguji validitas dan reliabilitas. Pertanyaannya adalah apakah soal-soal tersebut sudah sesuai dengan kisi-kisi yang diberikan BSNP, mampu mengukur kemampuan siswa dan dapat merepresentasikan UN yang kelak akan dihadapi siswa?

Membuat alat ukur yang baik tidaklah mudah, maka dari itu guru harus berusaha mengevaluasi soal yang telah dibuat. Terlebih lagi untuk mempersiapkan siswa dalam menghadapi UN, maka pembuatan soal uji coba UN harus diusahakan sebaik-baiknya supaya siswa benar-benar dapat mengambil pelajaran dari uji coba UN tersebut. Permasalahan lain yang diperoleh dari hasil observasi berkaitan dengan uji coba UN di Kabupaten Banjarnegara adalah siswa kurang serius dalam mengerjakan soal-soal tersebut. Sebagian besar guru SMA menyatakan uji coba UN seakan-akan disepelekan karena tidak masuk ke dalam penilaian apalagi menentukan kelulusan, akibatnya adalah nilai-nilai peserta didik buruk. Hal ini tentunya membuat guru menjadi cemas dan khawatir bila terjadi pada saat UN yang sebenarnya. Ada pula peserta didik yang mengalami kecemasan saat mengerjakan soal-soal *try out* UN. Dikarenakan pelaksanaan *try out* UN dengan perangkat soal yang dibuat oleh MGMP mendekati UN, maka tekanan semakin berat sehingga kecemasan tidak dapat dihindarkan.

Permasalahan kecemasan dan ketidakseriusan peserta tes salah satunya diantisipasi dengan menciptakan situasi yang kondusif untuk berkonsentrasi selama tes berlangsung. Untuk permasalahan alat ukur (tes), perlu dilakukan uji teoritis dengan melakukan analisis menggunakan Teori Tes Klasik (*Classical Test Theory/CTT*) dan/atau Teori Respon Butir (*Item Response Theory/IRT*). Metode-metode yang terdapat dalam CTT yaitu Thorndike, Polynomial, Binomial Lord dengan modifikasi Keats, Compound Binomial, dan Analisis Varians.

Oleh karena itu, penelitian ini dipandang perlu dilakukan agar guru-guru mata pelajaran ekonomi SMA di Kabupaten Banjarnegara pada khususnya dapat mengetahui kualitas soal uji coba UN yang telah dibuat dalam rangka persiapan menghadapi UN para peserta didiknya. Persiapan yang sebaik-baiknya menjelang UN merupakan salah satu bukti kesiapan guru untuk menghasilkan lulusan yang baik. Alasan pendukung lainnya, penelitian mengenai estimasi kesalahan pengukuran perangkat soal

uji coba UN terutama untuk mata pelajaran ekonomi SMA di Kabupaten Banjarnegara belum pernah dilakukan sebelum ini.

Tujuan dari penelitian ini adalah untuk (1) Mengetahui karakteristik perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara Tahun Ajaran 2014/2015. (2) Mengestimasi kesalahan pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara Tahun Ajaran 2014/2015. (3) Mengetahui metode yang paling akurat dalam mengestimasi kesalahan baku pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara Tahun Ajaran 2014/2015.

Metode Penelitian

Penelitian ini termasuk jenis penelitian deskriptif eksploratif. Termasuk penelitian deskriptif eksploratif karena akan mendeskripsikan data yang telah terkumpul dengan tujuan mengetahui karakteristik butir soal perangkat tes, serta mengetahui besarnya kesalahan pengukuran pada perangkat tes uji coba Ujian Nasional mata pelajaran ekonomi SMA Tahun Ajaran 2014/2015 di Kabupaten Banjarnegara dengan metode Thorndike, Binomial Lord Modifikasi Keats, Analisis Varians dan Teori Respon Butir, kemudian mencari metode yang paling akurat.

Tempat yang digunakan dalam penelitian ini adalah Kabupaten Banjarnegara Provinsi Jawa Tengah yang dilaksanakan pada bulan Februari sampai Maret 2015.

Obyek penelitian ini adalah lembar jawab siswa peserta uji coba Ujian Nasional mata pelajaran Ekonomi SMA yang dirancang oleh MGMP Ekonomi tahun ajaran 2014/2015 yang diperoleh dari 13 sekolah baik negeri maupun Swasta yang berjumlah sebanyak 809 lembar jawab sesuai dengan banyaknya jumlah peserta uji coba UN ekonomi pada tahun ajaran 2014/2015.

Teknik yang digunakan dalam penelitian ini adalah dokumentasi, dengan cara mengumpulkan lembar jawab siswa yang mengikuti uji coba Ujian Nasional ekonomi Tahun Ajaran 2014/2015 dari tiap-tiap sekolah yang menjadi tempat penelitian.

Teknik analisis data yang dilakukan dalam penelitian ini adalah analisis karakteristik perangkat soal uji coba Ujian Nasional mata pelajaran ekonomi SMA di Kabupaten Banjarnegara dilakukan dengan menggunakan Teori

Tes Klasik dan Teori Respon Butir. Kemudian dilanjutkan dengan estimasi kesalahan pengukuran dengan menggunakan Metode Thorndike, Binomial Lord Modifikasi Keats, Analisis Varians dan Teori Respon Butir.

Analisis karakteristik perangkat soal uji coba Ujian Nasional mata pelajaran ekonomi SMA di Kabupaten Banjarnegara terlebih dahulu dengan menggunakan Teori Tes Klasik. Analisis ini dilakukan untuk mengetahui karakteristik perangkat soal yang dibuat oleh MGMP Ekonomi SMA di Kabupaten Banjarnegara. 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) yaitu 0,3 sampai 0,7. Untuk kriteria daya beda, butir soal yang baik memiliki nilai $D \geq 0,3$ (Mardapi, 2012, p.188). Untuk distraktor atau pengecoh dikatakan berfungsi apabila nilai dari *Point Biser* pada setiap soal pilihan ganda minimum berindeks 0,1 dan bernilai positif sedangkan untuk jawaban salah berindeks negatif (Samsul Hadi:2011,p.3).

Analisis berdasarkan Teori Respon Butir memiliki tiga model logistik, yaitu model logistik satu parameter, model logistik dua parameter dan model logistik tiga parameter. Dari ketiga model logistik, maka kemudian ditentukan model logistik mana yang sesuai untuk perangkat soal uji coba Ujian Nasional mata pelajaran ekonomi SMA di Kabupaten Banjarnegara. Uji kecocokan model dilakukan menggunakan *software* BILOG-MG. 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 1. Kriteria Butir Soal yang Baik 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

Metode pertama yang digunakan untuk mengestimasi kesalahan pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara yaitu Metode Thorndike. Langkah-langkah yang ditempuh meliputi mencari reliabilitas (alpha cronbach) perangkat tes dari hasil analisis menggunakan Iteman. Selanjutnya membagi masing-masing perangkat tes menjadi dua bagian sama panjang, boleh berdasarkan genap-ganjil atau atas-bawah. Langkah selanjutnya mencari nilai rerata dan nilai varian pada masing-masing belahan, kemudian menghitung nilai varians kesalahan pada masing-masing belahan dengan menggunakan rumus $\sigma_{E1}^2 = \sigma_{x1}\sqrt{1-r_{xx}}$ dan $\sigma_{E2}^2 = \sigma_{x2}\sqrt{1-r_{xx}}$. Langkah selanjutnya mengestimasi kesalahan pengukuran dengan rumus yaitu $\sigma_E = (\sigma_{E1}^2 + \sigma_{E2}^2)^{1/2}$.

Metode kedua yang digunakan adalah metode Binomial Lord Modifikasi Keats. Langkah-langkah yang ditempuh adalah menentukan konstanta yang dilakukan dengan menghitung Alpha Cronbach dan KR-21 dari perangkat tes dan pembelahan perangkat soal berdasarkan tingkat kesukaran butir. Reliabilitas berdasarkan Alpha Cronbach diperoleh dari hasil analisis menggunakan Iteman. Perhitungan KR-21 dilakukan menggunakan formula $KR - 21 = \left(\frac{k}{k-1}\right) \left(1 - \frac{k\bar{p}(1-\bar{p})}{s_x^2}\right)$. Setelah Alpha Cronbach dan KR-21 diketahui, langkah selanjutnya adalah membelah perangkat soal menjadi empat bagian sama panjang berdasarkan tingkat kesukaran, boleh dari yang paling mudah atau yang paling sulit. Langkah selanjutnya melakukan estimasi kesalahan pengukuran perangkat tes menggunakan

$S_E = \left[\left(\frac{X(k-X)}{k-1}\right) \left(\frac{1-r_{xx'}}{1-r_{21}}\right)\right]^{1/2}$. Setelah itu hitung varians masing-masing bagian setelah itu hitung akar kuadrat dari penjumlahan keempat varians tersebut. Perhitungan dilakukan dengan menggunakan bantuan program komputer Ms Excel.

Metode estimasi yang ketiga yaitu metode Analisis Varians. Metode ini memandang skor soal dianggap sebagai desain faktorial dua jalan tanpa replikasi (Mardapi, 2008, p.53). Setiap butir soal dianggap sebagai *treatment* atau perlakuan. Langkah pertama yang dilakukan adalah mencari nilai *mean* kuadrat interaksi peserta tes dan butir soal dengan menggunakan rumus $MS_{Sxl} = \frac{\left(i - \frac{\sum X^2}{k} - \frac{\sum Y^2}{n}\right) + \frac{\sum i^2}{nk}}{(n-1)(k-1)}$. Pada metode Analisis Va-

rians, perangkat soal tidak perlu dibelah seperti pada metode Thorndike dan Binomial Lord Modifikasi Keats. Langkah selanjutnya adalah mengestimasi kesalahan baku pengukuran dengan menggunakan rumus $S_E = [k(MS_{Sxl})]^{1/2}$.

Metode estimasi kesalahan pengukuran yang keempat berdasarkan Teori Respon Butir dilakukan dengan bantuan Program Komputer *Bilog_MG*. Program tersebut akan menghasilkan nilai parameter butir soal, yaitu tingkat kesukaran, daya beda, dan *guessing*. Dari informasi tersebut maka akan diketahui model yang cocok dengan perangkat soal ini apakah 1PL, 2PL atau 3PL. Setelah diketahui model logistik yang cocok langkah selanjutnya yang dilakukan menghitung fungsi informasi perangkat tes dengan menggunakan rumus (Hambleton & Swaminathan: 1985, p.91) yaitu

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

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

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

Keterangan:

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

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

θ : 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)

disesuaikan dengan model logistik yang tepat pada tingkatan tetha $-3 \leq \theta \leq 3$ dan perubahan θ sebesar 0,1. Selanjutnya menghitung SEM dengan rumus $SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}$.

Hasil Penelitian dan Pembahasan

Hasil Penelitian

Penelitian ini dilakukan untuk membandingkan akurasi metode estimasi kesalahan pengukuran perangkat soal uji coba Ujian Nasional mata pelajaran Ekonomi SMA di Kabupaten Banjarnegara, Jawa Tengah. Metode yang digunakan dalam penelitian ini adalah metode Thorndike, Metode Binomial Lord

Modifikasi Keats, Analisis Varians dan Metode Teori Respon Butir. Perangkat soal yang digunakan adalah soal uji coba Ujian Nasional mata pelajaran Ekonomi SMA di Kabupaten Banjarnegara yang dibuat oleh Tim MGMP ekonomi di Kabupaten Banjarnegara. Berdasarkan penemuan di lapangan, diketahui bahwa tim MGMP Ekonomi SMA di Kabupaten Banjarnegara menyiapkan dua buah paket soal.

Total lembar jawab yang terkumpul yaitu sebanyak 809 lembar jawab siswa. Dari total lembar jawab sebanyak 809 tersebut terbagi menjadi dua paket. Untuk soal paket satu terdapat 404 lembar jawab siswa sedangkan soal paket dua terkumpul 405 lembar jawab siswa.

Dalam menganalisis perangkat soal uji coba UN mata pelajaran ekonomi berdasarkan teori tes klasik digunakan bantuan program komputer *MicroCat Iteman versi 3.0*. Informasi yang dapat diperoleh dari hasil perhitungan tersebut yaitu tingkat kesukaran, daya beda, dan keberfungsian distraktor. Hasil perhitungan telah dirangkum pada Tabel 2.

Tabel 2. Kesimpulan Kualitas Perangkat Soal Uji Coba Ujian Nasional Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara (Paket 1)

Kriteria	Jumlah butir	Persentase (%)
Baik (Semua parameter terpenuhi)	13	32,5
Kurang Baik (Satu atau dua parameter terpenuhi)	18	45,0
Tidak Baik (Semua parameter tidak terpenuhi)	9	22,5
	40	100

Berdasarkan Tabel 2 dapat disimpulkan bahwa perangkat soal yang digunakan untuk uji coba UN ekonomi di Kabupaten Banjarnegara (paket 1) masih tergolong kurang baik bila ditinjau dari segi parameternya. Reliabilitas yang dimiliki sebesar 0,667 dan tergolong kurang baik, karena kurang dari 0,7.

Berdasarkan Tabel 3 dapat disimpulkan bahwa perangkat soal yang digunakan untuk uji coba UN mata pelajaran Ekonomi di Kabupaten Banjarnegara tergolong baik bila ditinjau dari segi parameternya. Dikarenakan persentase soal yang baik sebesar 50% , maka bisa dikatakan perangkat soal kedua secara teori tes klasik lebih baik dibandingkan paket satu. Reliabilitas

soal paket dua sebesar 0,724 dan tergolong memiliki reliabilitas yang baik karena lebih dari 0,7.

Tabel 3. Kesimpulan Kualitas Perangkat Soal Uji Coba Ujian Nasional Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara (Paket 2)

Kriteria	Jumlah butir	Persentase (%)
Baik (Semua parameter terpenuhi)	20	50,0
Kurang Baik (Satu atau dua parameter terpenuhi)	13	32,5
Tidak Baik (Semua parameter tidak terpenuhi)	7	17,5
	40	100

Selanjutnya, analisis kuantitatif karakteristik berdasarkan Teori Respons Butir dilakukan untuk melihat model yang cocok dengan perangkat soal uji coba UN mata pelajaran Ekonomi di Kabupaten Banjarnegara. Analisis kuantitatif karakteristik soal berdasarkan Teori Respon Butir dilakukan dengan menggunakan program komputer *BilogMG*. Berikut hasil uji kecocokan model.

Tabel 4. Uji Kecocokan Model Perangkat Soal Uji Coba Ujian Nasional Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara (Paket 1)

Model	Jumlah Butir yang Cocok dengan Model
1 PL	25
2 PL	15
3 PL	11

Berdasarkan Tabel 4, maka dapat disimpulkan bahwa perangkat soal uji coba UN ekonomi paket 1 cocok dianalisis menggunakan teori respon butir 1 PL, karena jumlah butir yang cocok paling banyak dibandingkan model 2 PL dan 3 PL yaitu 25 butir. Selanjutnya perlu diuji untuk soal paket 2 cocok dengan model logistik 1 PL, 2 PL atau 3 PL. Berikut ini rangkuman uji kecocokan model untuk soal paket 2.

Mencermati banyak butir yang cocok dengan model logistik 1PL, maka dapat disimpulkan bahwa perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara (Paket 2) cocok dengan model 1 PL

menurut teori respons butir. Dengan demikian kedua paket soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara cocok dengan Teori Respons Butir dengan model logistik 1 PL.

Tabel 5. Uji Kecocokan Model Perangkat Soal Uji Coba Ujian Nasional Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara (Paket 2)

Model	Jumlah Butir yang Cocok dengan Model
1 PL	22
2 PL	14
3 PL	16

Berdasarkan keseluruhan hasil estimasi kesalahan pengukuran dengan menggunakan metode Thorndike, Binomial Lord Modifikasi Keats, Analisis Varians dan Teori Respons Butir diperoleh rangkuman sebagai berikut:

Tabel 5. Perbedaan Hasil Perhitungan Estimasi Kesalahan Pengukuran Perangkat Soal Uji Coba Ujian Nasional Ekonomi SMA di Kabupaten Banjarnegara

Metode Estimasi Kesalahan Pengukuran	Perangkat Soal Uji Coba UN Ekonomi (Paket 1)	Perangkat Soal Uji Coba UN Ekonomi (Paket 2)
Thorndike	1,8005	1,7673
Binomial Lord Modifikasi Keats	0,8742	0,8392
Analisis Varians	2,7873	2,8302
Teori Respon Butir	0,1990 s.d 1,0320	0,2005 s.d 1,0073

Berdasarkan pengaplikasian empat metode untuk mengestimasi kesalahan pengukuran perangkat soal uji coba UN Ekonomi, dapat diketahui bahwa metode yang menghasilkan nilai SEM terkecil (paling akurat) adalah metode Binomial Lord Modifikasi Keats yaitu sebesar 0,8742 untuk soal paket 1 dan 0,8392 untuk soal paket 2.

Pembahasan

Berdasarkan hasil analisis Teori Tes Klasik untuk mengetahui karakteristik Perangkat Soal Uji Coba UN Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara (Paket 1), diketahui bahwa terdapat 13 butir (32,5%) yang memenuhi kriteria ideal soal yang baik dari total soal sebanyak 40 butir soal. Butir soal

yang kurang atau tidak baik, disebabkan karena daya beda yang kurang dari 0,3 dan bahkan ada butir soal yang memiliki daya beda negatif (soal nomor 19 dan 37). Hal ini menandakan bahwa pada butir tersebut siswa dengan kemampuan atas menjawab salah sedangkan siswa dengan kemampuan bawah menjawab dengan benar. Dari parameter tingkat kesukaran, butir yang dianggap baik adalah butir soal dengan tingkat kesukaran menengah atau sedang, yaitu antara 0,3 – 0,7. Untuk butir dengan tingkat kesukaran < 0,3 maka tergolong soal sulit, seperti butir nomor 5, 8, 11, 12 dan lain sebagainya. Pada soal paket satu tidak ditemukan soal dengan tipe mudah atau yang memiliki tingkat kesukaran di atas 0,7. Jadi, soal yang dibuat oleh MGMP untuk soal paket satu terdiri atas soal sulit dan sedang/tenengah. Indeks reliabilitas perangkat soal paket satu sebesar 0,667 sehingga tergolong memiliki reliabilitas yang kurang baik karena kurang dari 0,7.

Estimasi kesalahan pengukuran menggunakan metode Thorndike menghasilkan nilai estimasi sebesar 1,8005. Nilai tersebut kemudian dimasukkan ke dalam rumus $x \pm z_{c,sE}$ untuk mengetahui rentang *true score* yang diperoleh masing-masing peserta tes. Berdasarkan hasil perhitungan kemampuan peserta tes, nilai sebenarnya yang diperoleh peserta tes yang menggunakan paket satu pada taraf kepercayaan 95% berada pada rentang nilai $x \pm 3,5299$.

Estimasi kesalahan pengukuran menggunakan metode Binomial Lord modifikasi Keats menghasilkan nilai estimasi sebesar 0,8742. Nilai tersebut kemudian dimasukkan ke dalam rumus $x \pm z_{c,sE}$ untuk mengetahui rentang *true score* yang diperoleh masing-masing peserta tes. Berdasarkan hasil perhitungan kemampuan peserta tes, nilai sebenarnya yang diperoleh peserta tes yang menggunakan paket satu pada taraf kepercayaan 95% berada pada rentang nilai $x \pm 1,7134$.

Estimasi kesalahan pengukuran menggunakan metode analisis varians menghasilkan nilai estimasi sebesar 2,7873. Nilai tersebut kemudian dimasukkan ke dalam rumus $x \pm z_{c,sE}$ untuk mengetahui rentang *true score* yang diperoleh masing-masing peserta tes. Berdasarkan hasil perhitungan kemampuan peserta tes, nilai sebenarnya yang diperoleh peserta tes yang menggunakan paket satu pada taraf kepercayaan 95% berada pada rentang nilai $x \pm 5,4631$.

Estimasi kesalahan pengukuran berdasarkan teori respon butir dapat dilihat dari

SEM tertinggi dan terendah yang diperoleh dari fungsi informasi dengan menyesuaikan model logistik yang cocok dengan perangkat soal ini yaitu model 1PL. Besarnya kesalahan pengukuran kemampuan peserta tes yang mengerjakan paket 1 berkisar antara 0,1990 s.d 1,0320.

Berdasarkan hasil perhitungan di atas, maka dapat disimpulkan bahwa metode Binomial Lord Modifikasi Keats merupakan metode yang paling akurat dibandingkan metode lain yang ada dalam penelitian ini, karena menghasilkan perhitungan SEM yang paling kecil. Hal ini karena Binomial Lord Modifikasi Keats mengikuti asumsi yang digunakan pada Metode Compound Binomial yaitu perangkat tes dibelah berdasarkan urutan tingkat kesukaran. Feldt, Steffen, dan Gupta (1985) *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 berdasarkan 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.

Berdasarkan hasil analisis Teori Tes Klasik untuk mengetahui karakteristik Perangkat Soal Uji Coba UN Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara (Paket 2), diketahui bahwa terdapat 20 butir (50%) yang memenuhi kriteria ideal soal yang baik dari total soal sebanyak 40 butir soal. Butir soal yang kurang atau tidak baik, disebabkan karena daya beda yang kurang dari 0,3 dan bahkan ada butir soal yang memiliki daya beda negatif (soal nomor 15). Hal ini menandakan bahwa pada butir tersebut siswa dengan kemampuan atas menjawab salah sedangkan siswa dengan kemampuan bawah menjawab dengan benar. Dari parameter tingkat kesukaran, butir yang dianggap baik adalah butir soal dengan tingkat kesukaran menengah atau sedang, yaitu antara 0,3 – 0,7. Untuk butir dengan tingkat kesukaran < 0,3 maka tergolong soal sulit, seperti butir nomor 4, 6, 10, 11 dan lain sebagainya. Pada soal paket dua juga terdapat soal dengan tingkat kesukaran di atas 0,7 sehingga tergolong soal yang mudah yaitu soal nomor 19, 23, 29 dan 37. Indeks reliabilitas perangkat soal kedua sebesar 0,724 sehingga tergolong memiliki reliabilitas yang baik.

Estimasi kesalahan pengukuran menggunakan metode thorndike menghasilkan nilai estimasi sebesar 1,7673. Nilai tersebut kemudi-

an dimasukkan ke dalam rumus $x \pm z_c SE$ untuk mengetahui rentang *true score* yang diperoleh masing-masing peserta tes. Berdasarkan hasil perhitungan kemampuan peserta tes, nilai sebenarnya yang diperoleh peserta tes yang menggunakan paket satu pada taraf kepercayaan 95% berada pada rentang nilai $x \pm 3,4639$.

Estimasi kesalahan pengukuran menggunakan metode Binomial Lord modifikasi Keats menghasilkan nilai estimasi sebesar 0,8392. Nilai tersebut kemudian dimasukkan ke dalam rumus $x \pm z_c SE$ untuk mengetahui rentang *true score* yang diperoleh masing-masing peserta tes. Berdasarkan hasil perhitungan kemampuan peserta tes, nilai sebenarnya yang diperoleh peserta tes yang menggunakan paket satu pada taraf kepercayaan 95% berada pada rentang nilai $x \pm 1,6448$.

Estimasi kesalahan pengukuran menggunakan metode analisis varians menghasilkan nilai estimasi sebesar 2,8302. Nilai tersebut kemudian dimasukkan ke dalam rumus $x \pm z_c SE$ untuk mengetahui rentang *true score* yang diperoleh masing-masing peserta tes. Berdasarkan hasil perhitungan kemampuan peserta tes, nilai sebenarnya yang diperoleh peserta tes yang menggunakan paket satu pada taraf kepercayaan 95% berada pada rentang nilai $x \pm 5,5472$.

Estimasi kesalahan pengukuran berdasarkan teori respon butir dapat dilihat dari SEM tertinggi dan terendah yang diperoleh dari fungsi informasi dengan menyesuaikan model logistik yang cocok dengan perangkat soal ini yaitu model 1PL. Besarnya kesalahan pengukuran kemampuan peserta tes yang mengerjakan paket 2 berkisar antara 0,2005 s.d 1,0073.

Berdasarkan hasil perhitungan di atas, maka dapat disimpulkan bahwa metode Binomial Lord Modifikasi Keats merupakan metode yang paling akurat dibandingkan metode lain yang ada dalam penelitian ini, karena menghasilkan perhitungan SEM yang paling kecil. Hal ini karena Binomial Lord Modifikasi Keats mengikuti asumsi yang digunakan pada Metode Compound Binomial yaitu perangkat tes dibelah berdasarkan urutan tingkat kesukaran. Feldt, Steffen, dan Gupta (1985) *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 berdasarkan 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.

Simpulan dan Saran

Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah disampaikan sebelumnya, maka simpulan yang dapat dibuat sebagai berikut. (1) Karakteristik perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara yang terdiri atas dua paket soal diketahui bahwa secara teori tes klasik soal paket dua lebih baik dibandingkan soal paket satu. Hal tersebut dapat dibuktikan dengan banyaknya soal yang memenuhi kriteria daya beda, tingkat kesukaran dan keberfungsian pengecoh pada soal paket dua lebih banyak dari pada soal paket satu. (2) Dalam mengestimasi kesalahan pengukuran perangkat soal uji coba UN SMA mata pelajaran ekonomi menggunakan empat metode, yaitu metode Thorndike, Binomial Lord Modifikasi Keats, Analisis Varians dan teori respon butir. Untuk soal paket satu, keempat metode tersebut secara berturut-turut menghasilkan estimasi kesalahan pengukuran sebesar 1,8005, 0,8742, 2,7873 dan 0,1990 s.d 1,0320. Untuk soal paket dua, besarnya estimasi kesalahan pengukuran yang dihasilkan oleh empat metode tersebut secara berturut-turut sebesar 1,7673, 0,8392, 2,8302 dan 0,2005 s.d 1,0073. (3) Metode yang paling akurat dalam mengestimasi kesalahan pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara adalah metode Binomial Lord modifikasi Keats baik pada soal paket satu maupun soal paket dua.

Saran

Berdasarkan hasil penelitian di atas, maka peneliti menyarankan: (1) Untuk pelaksanaan tes yang dibuat secara tim seperti Ujian Kenaikan Kelas dan uji coba UN hendaknya guru menggunakan bank soal yang baik, untuk meminimalisasi tingkat kesalahan pengukuran. (2) Hendaknya guru baik secara tim atau individu melakukan analisis estimasi kesalahan pengukuran secara berkelanjutan sebagai salah satu bentuk evaluasi hasil kerja dan juga dapat dijadikan bank soal. Jadi, evaluasi tidak hanya dengan melihat nilai masing-masing siswa atau rata-rata nilai yang dicapai para siswa dari waktu ke waktu tapi perlu juga mengevaluasi

soal yang telah dibuat. (3) Dinas pendidikan setempat hendaknya memfasilitasi guru-guru dengan memberikan pelatihan penyusunan soal yang baik dengan mengundang ahli.

Daftar Pustaka

- Azwar, Saifuddin. (2014). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Theory*. New York: CBS-College Publishing.
- Dali S. Naga. (1992). *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Feldt, L. S., Steffen, M., & Gupta, M.C. (1985). A comparison of five methods for estimating the standard error measurement at specific score level. *Applied Psychological Measurement*. Vol. 9. No. 4. Pp. 351-361.
- Hadi, S. (2014). *Lembar Kerja Teknik Analisis Data-Pascasarjana, Mengolah hasil analisis Bilog dengan SPSS*. Materi Kuliah Praktik Pengukuran dan Pengujian, Pascasarjana UNY.
- Hambleton, R. K. & Swaminathan H (1985). *Item response theory principles and applications*. New York: Sage Publications.
- Hambleton, R. K., Swaminathan H., & Rogers H. Jane. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications.
- Mardapi, D. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta: Nuha Litera.
- Miller, P. W. (2008). *Measurement and Teaching*. Muster: Partric W. Miller & Association.
- Mulyana. (2007). Karakteristik Soal Tes Masuk SMP Negeri di Kabupaten Bantul. *Jurnal Penelitian dan Evaluasi Pendidikan*, 10(2). dari <http://journal.uny.ac.id/index.php/jpep/article/view/1985>
- Reynolds, C. R., Livingston, R. B., & Wilson, V. (2010). *Measurement and Assessment in Education*. Upper Saddle River: Pearson.

