

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

Khotimah Marjiastuti

Program Pascasarjana Universitas Negeri Yogyakarta
khotimahmarjiastuti@ymail.com

Abstrak

Penelitian ini bertujuan untuk mengestimasi kesalahan pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA Tahun Ajaran 2014/2015 di Kabupaten Banjarnegara berdasarkan teori respon butir. Analisis data dilakukan berdasarkan respon peserta didik terhadap perangkat soal uji coba UN mata pelajaran ekonomi SMA tahun ajaran 2014/2015 sebanyak 809 lembar yang diperoleh dari 13 SMA negeri dan swasta. Metode yang digunakan yaitu Teori Respon Butir (Item Respond Theory). Hasil analisis menunjukkan perangkat soal cocok dengan model Teori Respon Butir 1 Parameter Logistik. Berdasarkan metode tersebut, nilai estimasi kesalahan pengukuran terkecil pada soal paket 1 terjadi pada siswa dengan $\theta=0$ dengan nilai sebesar 0,1990 sedangkan nilai estimasi terbesar terjadi pada siswa dengan $\theta=+3$ dengan nilai sebesar 1,0320. Nilai estimasi kesalahan pengukuran terkecil pada soal paket 2 terjadi pada siswa dengan $\theta=0$ dengan nilai sebesar 0,2005, sedangkan nilai estimasi terbesar terjadi pada siswa dengan $\theta=+3$ dengan nilai sebesar 1,0073. Hal ini menandakan bahwa perangkat soal yang disusun oleh MGMP ekonomi lebih cocok diberikan kepada siswa dengan kemampuan sedang.

Kata kunci: perangkat soal, kesalahan pengukuran, teori respon butir.

PENDAHULUAN

Lebih dari satu dekade lalu, para pemimpin ASEAN sepakat untuk membentuk sebuah pasar tunggal di kawasan Asia Tenggara pada akhir 2015 mendatang. Kesepakatan tersebut dilakukan agar daya saing ASEAN meningkat serta bisa menyaingi Tiongkok dan India untuk menarik investasi asing. Penanaman modal asing di wilayah ASEAN sangat dibutuhkan untuk meningkatkan lapangan pekerjaan serta meningkatkan kesejahteraan. Pembentukan pasar tunggal yang diberi nama Masyarakat Ekonomi ASEAN (MEA) memungkinkan suatu negara akan dengan mudahnya melakukan jual-beli barang dan jasa dengan negara lain di Asia Tenggara, sehingga secara otomatis persaingan akan semakin ketat.

Masyarakat Ekonomi ASEAN tidak hanya membuka arus perdagangan barang atau jasa, tetapi juga tenaga kerja profesional, seperti dokter, pengacara, akuntan, guru dan lain sebagainya. Dengan demikian, dengan adanya MEA maka akan semakin membuka peluang tenaga kerja asing untuk mengisi berbagai profesi di Indonesia. Sudah siapkah tenaga kerja Indonesia bersaing dengan negara lain di Asia Tenggara?

Salah satu tenaga profesional yang akan bersaing di era MEA adalah guru. Guru-guru di Indonesia diharapkan siap untuk bersaing dengan guru-guru asing. Oleh karena itu, tentu harus diikuti dengan kesiapan guru untuk meningkatkan kualitas, kemampuan

serta keterampilan di bidang pendidikan. Mengacu pada UU Nomor 14 Tahun 2005 menyatakan bahwa ada empat kompetensi yang harus dimiliki oleh guru, yaitu kompetensi pedagogik, kepribadian, profesional dan sosial. Keempat kompetensi tersebut memang idealnya dimiliki oleh guru di Indonesia. Salah satu di antara empat kompetensi yang ada yaitu kompetensi pedagogik. Kompetensi pedagogik menuntut guru untuk dapat memahami peserta didik dengan memanfaatkan prinsip-prinsip perkembangan kognitif dan kepribadian, perancangan pembelajaran, pelaksanaan pembelajaran yang kondusif, merancang dan melaksanakan evaluasi hasil belajar secara berkesinambungan dengan berbagai metode untuk kemudian sebagai acuan untuk menentukan tingkat ketuntasan belajar (*mastery learning*) dan memanfaatkan hasil evaluasi tersebut untuk perbaikan kualitas pembelajaran secara umum.

Salah satu kompetensi yang harus dimiliki guru adalah melakukan evaluasi hasil belajar secara berkesinambungan, namun kegiatan evaluasi dalam pendidikan tidak akan terlaksana tanpa adanya kegiatan pengukuran dan penilaian. Menurut Allen dan Yen (1979: 2), *measurement is the assigning of numbers to individuals in a systematic way as means of representing properties of the individuals*. Definisi tersebut dapat diartikan sebagai penetapan angka dengan cara yang sistematik untuk menyatakan keadaan individu. Pengukuran dilaksanakan dengan tujuan memberikan atribut berupa angka pada individu. Kegiatan pengukuran dilaksanakan dengan menggunakan instrumen tes dan/atau nontes. Penggunaan instrumen tes atau nontes disesuaikan dengan ranah yang hendak diukur. Allen dan Yen menyatakan bahwa *a test is a device for obtaining a sample of an individual's behavior* (Allen & Yen, 1979: 1). Kegiatan selanjutnya yaitu penilaian, yang merupakan pemberian label terhadap seseorang. Pemberian label dilakukan dengan memberikan kriteria atas nilai yang diperoleh. Nilai yang diperoleh dapat dikategorikan menjadi baik, cukup baik atau bahkan kurang baik.

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, dkk (2010) menyatakan bahwa sering orang mengandalkan kualitas soal hanya pada reliabilitasnya, namun di sisi lain bila diketahui, reliabilitas suatu tes tergantung pada tingkat kesalahan baku pengukuran (SEM). Kesalahan baku pengukuran erat kaitannya dengan koefisien reliabilitas suatu alat ukur. Miller (2008:93) mengemukakan, *SEM is a quantitative expression of the magnitude of error in a test score based on the test reliability. Selanjutnya, the reliability of the test scores decreases, the SEM increases. The greater the reliability of the test score, the smaller the SEM and the more confidence we have in the precision of the test score* (Reynolds, Livingston, dan Willson, 2010: 114). Pernyataan tersebut menjelaskan bahwa bila reliabilitas skor tes menurun maka kesalahan pengukuran yang terjadi justru akan meningkat. Reliabilitas yang tinggi akan menghasilkan kesalahan pengukuran yang

rendah. 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. Jadi, kualitas perangkat soal tidak hanya tergantung pada reliabilitas soalnya saja.

Kesalahan pengukuran dalam bidang pendidikan tidak dapat dihindari karena subjek dan objek yang terlibat yaitu manusia. Ada dua macam kesalahan pengukuran, yaitu kesalahan acak dan kesalahan sistematis (Mardapi, 2008: 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 merancukan 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.

Deteksi kesalahan baku pengukuran dapat dilakukan salah satunya dengan Teori Respon Butir. Hadirnya Teori Respons Butir (IRT) untuk menyempurnakan teori sebelumnya, yaitu Teori Tes Klasik (CTT). Kelemahan teori tes klasik salah satunya adalah tergantung pada karakteristik peserta tes (Mardapi, 2012: 189). Maksud dari pernyataan tersebut adalah bila sampel yang diberi soal kebetulan yang memiliki kemampuan tinggi, maka soal akan menjadi memiliki tingkat kesukaran yang rendah (indeks kesukaran tinggi). Sebaliknya bila soal diberikan kepada siswa dengan kemampuan rendah, maka soal akan menjadi memiliki tingkat kesukaran yang tinggi (indeks kesukaran rendah).

Model teori respon butir berdasarkan jumlah parameter butir ada tiga, yaitu 1-P, 2-P dan 3-P. Peluang menjawab benar suatu butir soal sebagai berikut (Mardapi, 2012: 203).

$$3 - P P_i(\theta) = c_i + (1 - c_i) \left(\frac{e^{1,7ai(\theta-b)}}{1+e^{1,7ai(\theta-b)}} \right)$$
$$2 - P P_i(\theta) = \left(\frac{e^{1,7ai(\theta-b)}}{1+e^{1,7ai(\theta-b)}} \right)$$
$$1 - P P_i(\theta) = \left(\frac{e^{1,7(\theta-b)}}{1+e^{1,7(\theta-b)}} \right)$$

Keterangan:

- a: daya pembeda
- b: tingkat kesukaran
- c: pseudo guessing
- e: eksponen
- θ : kemampuan

Parameter lain yang perlu diketahui dari teori respons butir yaitu fungsi informasi perangkat tes yang terdiri dari beberapa butir soal. Fungsi informasi dapat dihitung

dengan formula yang sesuai dengan model logistik perangkat tes yang akan dihitung nilai fungsi informasinya. Perhitungan nilai informasi perangkat tes untuk model logistik satu parameter sebagai berikut. (Hambleton, 1985: 91)

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

Perhitungan nilai informasi perangkat tes untuk model logistik dua parameter sebagai berikut

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

Perhitungan nilai informasi perangkat tes untuk model logistik tiga parameter sebagai berikut

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

Keterangan:

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

b_j : parameter indeks kesukaran butir ke-j

a_j : parameter daya beda butir ke-j

c_j : parameter *guessing* pada butir ke-j

e : bilangan transenden yang besarnya mendekati 2,718

D : nilai distribusi logistic besarnya 1,7

Setelah memperoleh fungsi informasi selanjutnya dapat dilakukan perhitungan SEM dengan Formula umum teori respons butir yang dibangun sebagai berikut.

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

Keterangan:

SEM : kesalahan baku pengukuran

θ : teta/tingkat kecerdasan peserta tes

$I(\theta)$: fungsi informasi pada nilai teta/tingkat kemampuan tertentu

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) dan lain-lain. 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).

Pada tahun 2015 UN tidak lagi menjadi penentu kelulusan yang utama bagi siswa, sejak Menteri Pendidikan dan Kebudayaan, Anies Baswedan, mengeluarkan keputusan untuk merevisi PP nomor 32 tahun 2013. Tujuan UN 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 penggeraan 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 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?

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 dengan sebaik-baiknya menjelang UN merupakan salah satu bukti kesiapan guru untuk menghasilkan lulusan yang baik. Di samping itu, untuk melakukan analisis estimasi kesalahan pengukuran guru harus memiliki kemampuan dan keterampilan mengoperasikan program komputer. Kemampuan dan keterampilan ini dapat menjadi bekal bagi guru dalam menghadapi persaingan di era MEA. 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.

METODE

Penelitian ini termasuk jenis penelitian deskriptif eksploratif. Objek penelitian adalah seluruh lembar jawab komputer (LJK) siswa peserta uji coba Ujian Nasional mata pelajaran Ekonomi SMA yang dirancang oleh MGMP Ekonomi tahun ajaran 2014/2015 dari 13 sekolah baik Negeri maupun Swasta atau sebanyak 809 LJK. Teknik yang digunakan dalam penelitian ini adalah dokumentasi. Estimasi kesalahan pengukuran berdasarkan Teori Respon Butir dilakukan dengan bantuan Program Komputer *Bilog_MG*.

Tabel 1. Objek Penelitian Estimasi Kesalahan Pengukuran Perangkat Soal Uji Coba Ujian Nasional Mata Pelajaran Ekonomi SMA di Kabupaten Banjarnegara

No	Nama Sekolah	Jumlah Peserta UN TA.2014/2015
1	SMAN 1 Banjarnegara	72
2	SMAN 1 Bawang	110
3	SMAN 1 Purwareja Klampok	81
4	SMA Muhammadiyah 4 Kalibening	28
5	SMAN 1 Sigaluh	58
6	SMA PGRI Purwareja Klampok	19
7	SMAN 1 Purwanegara	84
8	SMAN 1 Karangkobar	139
9	SMAN 1 Wanadadi	114
10	SMA Muhammadiyah 1 Banjarnegara	46
11	SMA Cokroaminoto Banjarnegara	20
12	SMAN 1 Batur	28
13	SMA Ma'arif Mandiraja	10
Total Peserta UN		809

HASIL DAN PEMBAHASAN

Langkah-langkah yang dilakukan untuk menganalisis SEM berdasarkan teori respon butir meliputi uji kecocokan model (1 PL, 2 PL atau 3 PL) dengan bantuan perangkat komputer BILOG MG, yang perlu dicermati adalah pada hasil perhitungan fase dua. Tentukan butir mana saja yang memenuhi semua kriteria. Setelah itu baru dapat dilanjutkan dengan mengestimasi kesalahan pengukuran perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara.

Uji Kecocokan Model

Untuk mengetahui perangkat soal tes uji coba UN cocok dengan model 1 PL, 2 PL atau 3 PL perlu dilakukan perhitungan dengan menggunakan BILOG MG. Ada tiga fase yang dihasilkan oleh *software* tersebut, fase satu menjelaskan informasi butir berdasarkan teori tes klasik, fase dua menjelaskan kualitas butir soal berdasarkan teori respon butir, sedangkan fase tiga memberikan informasi kemampuan (*ability*) masing-masing siswa dalam menjawab soal yang diberikan. Berikut ini rangkuman dari fase dua.

Tabel 2. Uji Kecocokan Model Perangkat Soal (Paket 1)

Model	Butir yang Cocok dengan Model	Jumlah Butir yang Cocok dengan Model
1 PL	1, 2, 5, 7, 9, 12, 13, 16, 17, 18, 20, 22, 23, 24, 25, 28, 29, 30, 32, 33, 35, 36, 28, 39, 40	25
2 PL	3, 6, 8, 9, 11, 14, 25, 26, 27, 28, 29, 31, 33, 34, 35	15
3 PL	3, 6, 7, 11, 14, 20, 26, 32, 33, 34, 35	11

Berdasarkan tabel di atas, 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. 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.

Tabel 3. Uji Kecocokan Model Perangkat Soal (Paket 2)

Model	Butir yang Cocok dengan Model	Jumlah Butir yang Cocok dengan Model
1 PL	4, 5, 6, 7, 8, 9, 11, 13, 16, 17, 18, 19, 21, 22, 24, 26, 28, 32, 35, 36, 38, 40	22
2 PL	1, 3, 5, 9, 13, 17, 19, 25, 29, 30, 32, 33, 37, 38	14
3 PL	2, 3, 5, 9, 17, 19, 21, 24, 25, 26, 29, 32, 33, 35, 38, 40	16

Mencermati banyak butir yang cocok dengan model logistik 1PL, maka dapat disimpulkan bahwa perangkat soal uji coba UN mata pelajaran ekonomi SMA paket 2 cocok dengan model 1 PL menurut teori respon butir.

Estimasi Kesalahan Pengukuran

Setelah diketahui model yang cocok yaitu model IRT 1 PL, selanjutnya adalah mencari fungsi informasi butir dengan menggunakan formula 3 dan 4 dengan menghitung mulai dari tetha -3 s.d +3 selisih antar tetha 0,1. Setelah diketahui fungsi informasi selanjutnya dapat dilakukan perhitungan SEM dengan formula 7. Setelah itu dapat diketahui bahwa SEM untuk perangkat soal pertama berkisar antara 0,1990 sampai 1,0320. SEM terkecil berada pada siswa dengan kemampuan $\theta = 0$ dan SEM tertinggi berada pada siswa dengan kemampuan $\theta = +3$. Sedangkan SEM untuk perangkat soal kedua berkisar antara 0,2005-1,0073 (SEM terendah sampai SEM tertinggi). SEM terkecil berada pada siswa dengan kemampuan $\theta = 0$ dan SEM tertinggi berada pada siswa dengan kemampuan $\theta = +3$.

SIMPULAN

Berdasarkan uji kecocokan model, kedua perangkat soal uji coba UN mata pelajaran ekonomi SMA di Kabupaten Banjarnegara cocok dengan model IRT 1 parameter logistik. Besarnya estimasi kesalahan pengukuran (SEM) untuk soal paket satu berdasarkan teori respon butir sebesar 0,1990-1,0320 sedangkan untuk paket dua 0,2005-1,0073. Kedua perangkat soal memiliki kesamaan yaitu SEM terendah terjadi pada siswa dengan kemampuan $\theta=0$ dan SEM tertinggi terjadi pada siswa dengan $\theta=+3$. Hal ini menandakan bahwa perangkat soal yang disusun oleh MGMP ekonomi lebih cocok diberikan kepada siswa dengan kemampuan sedang.

Saran yang dapat peneliti berikan sebagai berikut:

1. Untuk pelaksanaan tes yang dibuat secara tim seperti Ujian Kenaikan Kelas dan uji coba UN hendaknya guru menggunakan soal yang telah valid dan reliabel, 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. Jadi, evaluasi tidak hanya dengan melihat nilai atau rata-rata nilai yang dicapai 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

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Wadsworth
- Hambleton, R.K. & Swaminathan H. (1985). *Item Response Theory Principles and Applications*. New York: Springer.
- Mardapi, Djemari. (2008). *Teknik penyusunan instrumen tes & non tes*. Yogyakarta: Mitra Cendikia.
- Mardapi, Djemari. (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Yogyakarta: Nuha Litera.
- Miller, P. W. (2008). *Measurement and teaching*. Muster: Partric W. Miller & Association.
- Tighe, J., McManus, I. C., Dewhurst, N. G., Chis, L., & Mucklow, J. (2010) The Standard Error of Measurement is a more Appropriate Measure of Quality for Postgraduate Medical Assessments Than is Reliability: An Analysis of MRCP (UK) Examinations. *BMC Medical Education*, 10:40.