

**ESTIMASI KEMAMPUAN SISWA DALAM UJIAN NASIONAL
MENGUNAKAN METODE BAYES**

***ESTIMATION OF STUDENT ABILITY IN NATIONAL EXAMINATION
USING BAYESIAN METHOD***

Eviana Hikamudin

Peneliti pada Pusat Penilaian Pendidikan, Balitbang, Kemendikbud
Jalan Gunung Sahari Raya, Jakarta Pusat
e-mail: eviana.hikamudin@kemdikbud.go.id

Diterima: 15 Mei 2017; dikembalikan untuk direvisi: 22 Juni 2017; disetujui: 25 Juli 2017

ABSTRAK

Tujuan penelitian ini adalah untuk mengukur estimasi kemampuan siswa SMA Program IPA dalam Ujian Nasional (UN) mata pelajaran Matematika dengan menggunakan metode Rerata Bayes. Metode penelitian yang digunakan adalah metode Deskriptif Kuantitatif ex post-facto. Data dalam penelitian ini adalah hasil UN berupa respon jawaban siswa SMA Program IPA mata pelajaran Matematika tahun pelajaran 2015/2016. Sampel dipilih secara acak sebanyak 1200 responden dan jumlah butir sebanyak 40 butir. Hasil penelitian menunjukkan bahwa sebanyak 12,08% kemampuan siswa berada pada kategori tinggi, 74,28% berada pada kategori sedang, dan 13,65% berada pada kategori rendah. Simpulan dari penelitian ini adalah dengan menggunakan metode Bayes diperoleh hasil estimasi kemampuan siswa SMA IPA dalam UN mata pelajaran matematika tahun pelajaran 2015/2016 sebagian besar berada pada level sedang (rerata). Metode Bayes cukup akurat digunakan untuk mengestimasi kemampuan siswa, sehingga dapat menghasilkan informasi yang obyektif untuk meningkatkan mutu pembelajaran.

Kata kunci: estimasi parameter, kemampuan siswa, metode Rerata Bayes

ABSTRACT

The purpose is student ability measuring estimation in Senior High School of Science Program in Mathematics at National Examination (UN) using Bayesian method. The method of research is Quantitative Descriptive. The data is result of UN for Mathematics academic years 2015/2016. Samples have got by randomize about 1200 students and 40 items test.. The distribution of student's ability are: 12,08% in category high, 74,28% in medium category, and 13,65% in low category level. The conclusion is student ability in Senior High School of Science Program in Mathematics at UN academic years 2015/2016 which obtain of Bayesian Method tent to be average. Bayesian Method is accurate to estimate student ability and make objective informations to improve learning quality.

Keywords: parameter estimation, student's ability, Bayesian method

PENDAHULUAN

Secara garis besar terdapat tiga bahasan utama di dalam kajian uji tes dan pengukuran pendidikan. Naga (1992: iii) menyebutkan tiga bahasan utama tersebut adalah konstruksi uji tes atau alat ukur, penyelenggaraan uji tes, dan penskoran pada uji tes. Ketiga bahasan tersebut merupakan satu kesatuan komponen yang saling mempengaruhi dalam menghasilkan hasil pengukuran pendidikan yang terukur dan objektif.

Salah satu bahasan yang menarik untuk dikaji adalah penskoran hasil ujian nasional (UN). Dalam perspektif evaluasi pendidikan, UN merupakan salah satu bentuk evaluasi untuk mengukur keberhasilan pendidikan secara nasional. Hasil UN dapat dijadikan sebagai acuan dalam menentukan peta mutu pendidikan dalam suatu satuan pendidikan atau wilayah tertentu. Peta mutu pendidikan dapat diukur melalui capaian standar minimal kompetensi peserta didik dalam UN. Sulistyio (2007) menyebutkan bahwa capaian standar minimal UN bukan hanya merupakan tantangan bagi siswa, tetapi juga merupakan *washback effect* yang positif bagi semua pihak, terutama guru dan orang tua. Guru diharapkan dapat menjalankan fungsi pengajaran dengan lebih bersungguh-sungguh sehingga dapat mengantarkan siswa dalam kegiatan pembelajaran. Pendapat lain yang memperkuat pentingnya UN untuk dilaksanakan dikemukakan oleh Hartanto (2015) yaitu bahwa UN berdasarkan filosofi pendidikan yang berpedoman pada falsafah Pancasila menunjukkan adanya kepentingan menyelamatkan ciri khas dan budaya bangsa Indonesia, disamping itu adanya keterkaitan dan tidak terpisahkannya antara UN dengan tujuan pendidikan nasional.

Selama ini hasil UN yang diberikan kepada siswa adalah berupa nilai agregat dari total skor jawaban butir yang dijawab benar oleh responden (siswa). Jadi, sesungguhnya nilai tersebut didasarkan pada skor mentah (*raw score*) dari jumlah jawaban yang benar. Cara penskoran seperti ini belum dapat menggambarkan kemampuan responden sesungguhnya karena belum memperhitungkan perbedaan taraf sukar butir pada setiap soal yang dijawab siswa. Sebagai contoh dua orang siswa yang mendapat skor yang sama, belum tentu menjawab pada butir soal yang sama. Padahal setiap butir memiliki taraf sukar (b) yang berbeda-beda, sehingga jika salah satu siswa menjawab butir soal yang lebih sulit seharusnya mendapatkan nilai yang berbeda dengan siswa lainnya yang mendapat butir soal yang lebih mudah. Di sinilah terjadi kelemahan dalam penskoran UN karena pendekatan penskorannya masih menggunakan pendekatan Teori Klasik (*Classical Theory*).

UN adalah kegiatan pengukuran dan penilaian pencapaian kompetensi lulusan siswa secara nasional pada mata pelajaran tertentu. Dengan demikian, hasil UN merupakan sebuah laporan yang seharusnya dapat menggambarkan kemampuan siswa yang sebenarnya. Berdasarkan ketentuan yang berlaku saat ini, hasil UN digunakan untuk 1) pemetaan mutu program dan/atau satuan pendidikan; 2) pertimbangan seleksi masuk jenjang pendidikan berikutnya; dan 3) pertimbangan dalam pembinaan dan pemberian bantuan pada satuan pendidikan dalam upayanya untuk meningkatkan mutu pendidikan.

Berdasarkan latar belakang yang telah diuraikan di atas, rumusan masalahnya adalah

sebagai berikut: (1) bagaimana hasil estimasi kemampuan siswa SMA IPA dalam UN tahun 2015/2016 mata pelajaran matematika dengan menggunakan Metode Bayes?, (2) bagaimana akurasi Metode Bayes dalam mengestimasi kemampuan siswa?

Penelitian ini bertujuan untuk: (1) mengukur hasil estimasi kemampuan siswa SMA IPA dalam UN tahun 2015/2016 mata pelajaran matematika dengan menggunakan Metode Bayes. (2) mengukur akurasi Metode Bayes dalam mengestimasi kemampuan siswa.

KAJIAN PUSTAKA

Dalam pengukuran pendidikan, selain pendekatan Teori Klasik terdapat pendekatan Teori Modern (*Item Response Theory - IRT*). Terdapat perbedaan mendasar antara Teori Klasik dan Teori Modern. Salah satu perbedaannya adalah bahwa dalam Teori Klasik parameter butir yang diukur sangat tergantung pada kemampuan responden. Bila responden berkemampuan tinggi maka butir tes akan tampak mudah dan memiliki taraf sukar rendah. Sebaliknya, jika responden berkemampuan rendah maka butir tes akan tampak sulit dan taraf sukarnya tinggi. Dengan demikian, penerapan Teori Klasik dalam pengukuran faktor perbedaan kemampuan dan karakteristik tes merupakan keterbatasan dan kelemahan. Keterbatasan dan kelemahan yang terdapat dalam Teori Klasik dapat ditutupi oleh munculnya Teori Modern (IRT) di mana parameter butir yang diukur tidak tergantung pada kemampuan responden.

Dalam Teori Modern (IRT) pengukuran parameter dilakukan dengan suatu cara yang disebut dengan estimasi. Berdasarkan data

yang telah diperoleh melalui pengukuran, selanjutnya dilakukan estimasi terhadap parameter yang mencakup estimasi parameter kemampuan responden dan estimasi parameter butir. Estimasi parameter kemampuan responden dilakukan melalui pengukuran satu responden yang merespon sejumlah butir, sedangkan estimasi parameter butir dilakukan melalui satu butir yang direspon oleh sejumlah responden. Estimasi parameter merupakan sebuah cara yang efektif untuk mengetahui dan menentukan nilai parameter yang mendekati keadaan sebenarnya. Diana dan Soehardjoepri (2016) menjelaskan bahwa estimasi adalah suatu metode dimana kita dapat memperkirakan nilai dari suatu populasi dengan menggunakan nilai dari sampel. Terdapat dua jenis estimasi yaitu estimasi titik dan estimasi interval. Estimasi titik adalah nilai yang berfungsi untuk suatu pendugaan dari parameter populasi, sedangkan estimasi interval adalah interval yang menyatakan keberadaan dari suatu parameter populasi.

Dalam melakukan pengestimasi parameter kemampuan terdapat beberapa persyaratan yang perlu dipenuhi. Naga (1992:247) menyebutkan beberapa persyaratan yang perlu dipenuhi oleh butir tes dan peserta juga oleh gabungan di antara keduanya pada IRT adalah 1) Model responsi butir yang akan digunakan perlu ditentukan terlebih dahulu, 2) Banyaknya parameter yang akan diestimasi perlu ditentukan, 3) Butir yang akan digunakan dalam ujites perlu memenuhi syarat unidimensi dan peserta dalam ujites tersebut perlu memenuhi syarat independensi lokal, 4) karakteristik butir pada ujites harus memenuhi syarat invariansi terhadap butir dan terhadap peserta.

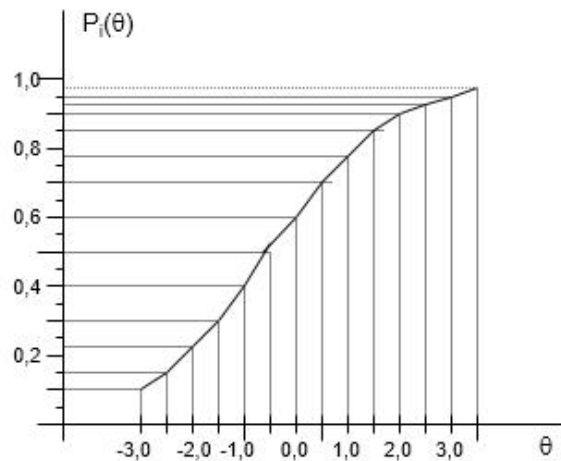
Proses pengestimasi parameter dalam IRT tidak bisa dilepaskan dari model karakteristik butir yang digunakan. Di dalam model responsi butir atau karakteristik butir terdapat tiga unsur pokok yang saling berkaitan, yaitu unsur butir, unsur peserta, dan unsur responsi atau jawaban peserta terhadap butir. Ketiga unsur tersebut saling berhubungan, sehingga menghasilkan sebuah fungsi matematik atau lengkungan responsi atau lengkungan karakteristik butir. Di dalam hubungan ini Naga (1992) menjelaskan bahwa ciri peserta dinyatakan melalui parameter ciri peserta, ciri butir dinyatakan melalui tiga parameter butir, yaitu daya beda (*a*), taraf sukar (*b*), dan kebetulan menjawab dengan benar (*c*), sedangkan ciri responsi atau jawaban peserta terhadap butir dinyatakan dalam bentuk probabilitas jawaban benar. Dengan demikian, Naga (1992:166) menuliskan hubungan tersebut sebagai:

$$P(\theta) = f(\theta, a, b, c) \dots\dots\dots \text{rumus (1)}$$

Model responsi butir atau karakteristik butir merupakan hal yang sangat penting diperhatikan dalam menerapkan IRT. Karakteristik butir merupakan fungsi matematika yang menggambarkan probabilitas menjawab benar dari responden terhadap butir sesuai dengan kemampuan (*ability*) dalam menjawab butir tersebut. Hambleton, dkk (1991:12) menyebutkan, “An item characteristic function or item characteristic curve (ICC) is a mathematical expression that relates the probability of success on an item to the ability measured by the test and the characteristics of the item.”

Menurut para ahli di bidang pengukuran model responsi yang pada umumnya diterima

adalah model ojaif normal. Model ojaif normal merupakan model karakteristik butir yang menggunakan lengkungan ojaif normal dalam membangun model karakteristik butir. Model ojaif normal digambarkan sebagai berikut:



Grafik 1. Model Ojaif Normal

Sumbu horisontal pada grafik 1 menunjukkan kemampuan (*ability*) peserta tes dan sumbu vertikalnya menunjukkan probabilitas (peluang) peserta tes menjawab soal dengan benar. Model ojaif normal sebagaimana di digambarkan pada Grafik 1 terbentuk berdasarkan persamaan probabilitas distribusi normal berikut.

$$\begin{aligned} Q_i(\theta) &= 1 - P_i(\theta) \\ &= \int_{(\theta-b_i)}^{\infty} n(\theta;0,1) d\theta \\ &= \int_{(\theta-b_i)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2} d\theta \\ &= 1 - \Phi(\theta-b_i) \end{aligned}$$

Menurut Naga (1992:192) model ojaif normal dianggap paling memadai untuk banyak hal di bidang pendidikan termasuk ujites tetapi model ini memerlukan perhitungan yang cukup rumit. Oleh karena itu, para ahli telah menemukan model lain

yang menyerupai model ojaif normal, tetapi dengan perhitungan yang lebih sederhana serta memiliki lebih banyak fleksibilitas. Model alternatif ini dikenal dengan nama Model Logistik. Azwar (2015:141) menjelaskan salah satu sebab mengapa model logistik lebih banyak digunakan adalah karena lebih sederhana dalam prosedur komputasinya dibandingkan dengan prosedur komputasi pada model ojaif normal.

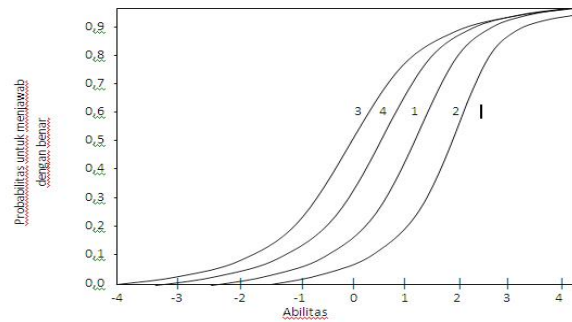
Model logistik terdiri dari model satu parameter (L1P), model dua parameter (L2P), dan model tiga parameter (L3P). Penggunaan nama-nama tersebut sesuai dengan banyaknya parameter yang digunakan. Model L1P hanya menggunakan satu parameter, yaitu parameter taraf sukar butir (b). Model L2P menggunakan dua parameter, yaitu selain menggunakan parameter taraf sukar butir (b) dan parameter daya beda butir (a). Model L3P menggunakan tiga parameter, yaitu parameter taraf sukar butir (b), parameter daya beda (a), dan parameter kebetulan menjawab dengan benar (c). Salah satu model karakteristik butir yang sering digunakan adalah model karakteristik logistik 1 parameter (L1P). Dalam model ini parameter yang diukur adalah parameter kemampuan responden (θ) dan parameter taraf sukar butir (b). Model logistik ini dipandang fleksibel dalam penggunaannya karena tidak memerlukan perhitungan yang rumit. Berikut adalah persamaan probabilitas model karakteristik logistik 1 parameter.

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} \text{ atau}$$

$$P_i(\theta) = \frac{1}{1 + e^{-D(\theta - b_i)}}$$

Berdasarkan persamaan di atas, grafik model karakteristiknya seperti digambarkan

di bawah ini.



Grafik 2. Model Logistik 1 Parameter

Dalam teori pengukuran terdapat banyak model responsi yang dapat dipilih untuk menggambarkan karakteristik butir. Namun, model yang dipilih seharusnya merupakan model yang dapat merepresentasikan keadaan responden yang sebenarnya. Hambleton dkk (1991:12) menjelaskan, “*The choice of model is up to the user; but this choice involves assumptions about the data that can be verified later by examining how well the model explains the observed test results.*”

Terdapat beberapa metode yang dapat digunakan dalam melakukan estimasi parameter. Setiap metode memiliki karakteristik masing-masing, baik dalam rumusan, prosedur, maupun kerumitan dalam penganalisisan. Di samping itu, setiap metode yang digunakan memiliki kekurangan dan kelebihan masing-masing. Tidak jarang terjadi bahwa suatu model yang digunakan cermat dalam perhitungan tetapi rumit, sehingga sukar untuk diterapkan. Sebaliknya, terkadang model tertentu sederhana dalam penerapannya tetapi lemah dalam kecermatannya.

Setiap metode estimasi yang digunakan memiliki keunggulan dan kelemahan masing-masing. Salah satu indikator yang dapat

digunakan adalah dengan membandingkan estimasi kekeliruan bakunya. Semakin kecil estimasi kekeliruan bakunya maka semakin akurat metode tersebut atau sebaliknya semakin besar estimasi kekeliruan bakunya maka semakin lemah akurasi.

Berkaitan dengan akurasi hasil estimasi, Embretson dan Reise (2000:179) menjelaskan, “*The EAP (Bayesian) estimators has minimum mean square error over the population of ability*”. Pendapat ini menyatakan bahwa variansi estimasi metode Bayes relatif kecil. Hal ini mengandung arti bahwa metode Bayes secara umum menghasilkan deviasi standar yang lebih kecil sehingga lebih akurat karena semakin kecilnya nilai *standard error*.

Metode estimasi Rerata Bayes seringkali digunakan karena memiliki keunggulan dalam perhitungan akurasi. Estimasi Rerata Bayes didasarkan pada distribusi diskrit. Metode estimasi Bayes menggunakan pendekatan “*quadrature weight*”, yaitu menghitung bobot hasil estimasi (*aposteriori*) berdasarkan asumsi dasar (*prior*) yang berupa distribusi tertentu pada populasi. Salah satu keunggulan metode ini adalah dapat mengestimasi semua respon yang benar semua ataupun yang salah semua. Secara matematis, de Ayala (2009:357) menuliskan metode Bayes sebagai probabilitas dengan rumus:

$$p(x) = \sum^R p(x / X_r, b) A(X_r) \dots\dots\dots \text{rumus (2)},$$

di mana:

- \underline{x} : vektor respon (respon terhadap butir),
- X_r : *quadrature point*,
- $A(X_r)$: *quadrature weight* , dan
- b : taraf sukar butir.

Yendra dan Noviadi (2015) menjelaskan bahwa Metode Bayes merupakan metode yang telah mendapatkan tempat pada para peneliti dalam mengestimasi parameter suatu distribusi. Metode ini sangat baik digunakan terutama bagi fungsi distribusi yang sangat rumit, atau dengan kata lain parameter yang dimiliki oleh fungsi distribusi tersebut lebih dari dua parameter. Keunggulan lainnya dari Metode Bayes adalah sebagaimana disebutkan oleh Reskianti, Nurtiti, dan Nasrah (2014) adalah memiliki distribusi awal (*prior*) yang kemudian digabungkan dengan data sampel sehingga menghasilkan data posterior yang cukup akurat untuk mengestimasi parameter populasi.

Berdasarkan uraian di atas maka dilakukan penelitian untuk mengukur kemampuan siswa dalam menjawab soal-soal UN melalui pendekatan IRT. Diharapkan dengan penelitian ini diperoleh bahan rujukan untuk melakukan penskoran hasil UN dengan pendekatan IRT, sehingga diperoleh hasil UN yang dapat menggambarkan kemampuan siswa secara objektif.

METODE PENELITIAN

Data yang diolah dalam penelitian ini adalah respon (jawaban) siswa SMA Program IPA dalam UN mata pelajaran Matematika tahun pelajaran 2015/2016. Sampel dipilih secara acak sebanyak 1200 responden dan jumlah butir sebanyak 40 butir. Banyaknya sampel butir yang digunakan didasarkan pada jumlah butir soal UN mata pelajaran matematika yang berjumlah 40 butir, sedangkan banyaknya responden yang dipilih dalam penelitian ini didasarkan pada pertimbangan stabilitas data penelitian. Suwanto (2005) menyebutkan dalam hasil

penelitiannya bahwa ada pengaruh ukuran sampel terhadap hasil estimasi parameter. Urutan ukuran sampel yang terbaik digunakan dalam mengestimasi adalah $N=1000$, $N=500$, dan $N=200$. Dalam penelitian ini sampel yang dipilih adalah sebanyak 1200 responden. Dengan demikian dapat dianggap bahwa penentuan jumlah sampel (responden) dalam penelitian sudah memadai.

Metode yang digunakan dalam penelitian ini adalah metode Deskriptif Kuantitatif *ex post-facto*. Sevilla dkk (1993:71) menjelaskan bahwa metode deskriptif digunakan untuk mengumpulkan informasi tentang keadaan-keadaan nyata yang telah berlangsung. Sesuai dengan tujuannya, penelitian yang dilakukan ini untuk mengukur parameter kemampuan siswa dalam UN tahun 2015/2016 melalui model karakteristik logistik 1 parameter dengan menggunakan metode Bayes. Data yang diolah dan dianalisis merupakan data yang sudah tersedia sebelumnya yaitu data hasil UN.

Populasi penelitian ini terdiri dari dua jenis, yaitu seluruh peserta UN SMA Tahun Pelajaran 2015/2016 dan butir soal UN SMA Tahun Pelajaran 2015/2016 mata pelajaran Matematika. Sampel penelitiannya adalah responden peserta UN SMA Tahun Pelajaran 2015/2016 sebanyak 1200 orang dan respon butir soal UN SMA Tahun Pelajaran 2015/2016 mata pelajaran Matematika sebanyak 40 butir.

Pemilihan banyaknya sampel sebanyak 1200 orang dan banyak butir sebanyak 40 butir didasarkan pada pertimbangan bahwa dengan jumlah sampel-sampel tersebut dianggap sudah memadai untuk tercapainya kestabilan dalam estimasi. Hal ini sesuai dengan hasil penelitian yang dilakukan oleh Retnawati

(2015) yang menyatakan bahwa estimasi yang melibatkan 25 dan 30 butir, serta peserta lebih dari 1.000 akan memperoleh hasil yang baik. Hal ini diperkuat oleh penelitian yang dilakukan oleh Setiawan (2009) yang salah satu hasilnya menyebutkan bahwa makin besar ukuran sampel akan makin kecil variansi estimasinya, yang artinya akan semakin akurat.

Data yang diolah dalam penelitian ini adalah data sekunder hasil UN yang diperoleh dari Pusat Penilaian Pendidikan, Balitbang Kemdikbud. Data yang digunakan dalam penelitian ini diambil secara acak (random) dari bank data hasil Ujian Nasional Sekolah Menengah Atas (SMA) Mata Pelajaran Matematika Tahun Pelajaran 2015/2016. Responden dipilih secara acak dan diperoleh sebanyak 1500 responden. Sedangkan butir soal yang dipilih adalah sebanyak 40 butir soal sesuai dengan paket soal yang dikerjakan oleh responden. Data yang telah dikumpulkan tersebut kemudian disusun ke dalam tabulasi data yang diperlukan.

Pengolahan dan analisis data dilakukan dengan menggunakan program *Bilog MG versi 3* untuk mengukur estimasi parameter kemampuan responden (θ) dan taraf sukar butir (b). Mathilda (2003) menguraikan bahwa keunggulan yang dimiliki oleh program *Bilog MG* adalah dapat mengukur estimasi dengan menggunakan beberapa metode estimasi yang salah satunya adalah Bayes. Proses estimasi dilakukan dengan cara replikasi sebanyak 5 kali. Tujuan dilakukannya replikasi ini adalah untuk melihat konsistensi hasil estimasi sehingga diperoleh hasil estimasi yang mendekati nilai sebenarnya.

Pengolahan dan analisis data dilakukan dengan dengan cara menyusun data input

berupa respon (jawaban) siswa. Selanjutnya dilakukan uji persyaratan independensi lokal antarresponden terhadap 1200 responden dengan menggunakan program *LOCINDEP*. Hikamudin (2016:50) menjelaskan bahwa program *LOCINDEP* merupakan program aplikasi yang khusus dirancang untuk melakukan uji independensi lokal dengan cara probabilitas.

Independensi lokal merupakan salah satu syarat yang harus dipenuhi dalam melakukan pengukuran dengan menggunakan IRT. Sesuai dengan definisinya, independensi lokal secara matematis dirumuskan sebagai berikut:

$$P(X_1 \cap X_2 \cap \dots \cap X_n) = \prod_{j=1}^n P(X_j)$$

di mana:

N : banyaknya butir ujites

X_j : skor peserta di dalam subpopulasi

Untuk menunjukkan terjadinya independensi lokal, dilakukan perhitungan probabilitas dengan persamaan $\pi_i = \dots$

Butir

Responden ke-1 : 1 0 1 1 0 1 1 0 0 0
 Responden ke-2 : 1 0 0 0 1 1 0 0 1 0

matriks proporsi.

		Responden ke-2		
		1	0	
Responden ke-1	1	p11	p10	x1
	0	p01	p00	x2
		y1	y2	xy

Keterangan:

1 : sekor menjawab benar

0 : sekor menjawab salah

p : proporsi

x_i : jumlah proporsi pada baris ke-i

y_i : jumlah proporsi pada kolom ke-i

Independensi lokal terjadi apabila memenuhi: $p_{11} = x_1y_1$, $p_{10} = x_1y_0$, $p_{01} = x_0y_1$, dan $p_{00} = x_0y_0$.

Untuk melakukan pengolahan data estimasi, terlebih dahulu dipilih subkelompok sampel secara acak (random) sebanyak 4 kali yang masing-masing berjumlah 300 responden dari 1200 responden. Kemudian menyusun *syntax* dalam program *Bilog MG* untuk mengukur estimasi parameter dengan menggunakan metode Bayes, menjalankan (*running*) program *Bilog MG* untuk melakukan estimasi parameter kemampuan siswa dengan replikasi sebanyak 4 kali, dan menganalisis hasil estimasi kemampuan siswa.

HASIL DAN PEMBAHASAN

Pengujian Independensi Lokal

Berdasarkan hasil pengujian independensi lokal dengan menggunakan program *LOCINDEP*, diperoleh data bahwa lebih dari 94% hasil pengujian memenuhi kriteria independensi lokal. Dengan demikian dapat diasumsikan bahwa sampel yang digunakan dalam penelitian ini memenuhi syarat untuk pengujian dengan menggunakan IRT.

Estimasi Parameter Kemampuan Responden

Estimasi parameter kemampuan responden dilakukan dengan cara replikasi sebanyak 4 kali yaitu memilih subkelompok sampel secara acak sebanyak 4 kali yang masing-masing berjumlah 300 responden, kemudian subkelompok yang terpilih masing-masing diestimasi kemampuannya dengan Metode Bayes. Dari pemilihan sampel tersebut terdapat beberapa responden yang terambil lebih dari 1 kali ke dalam

subkelompok sampel yang berbeda. Dengan demikian, di antara beberapa subkelompok sampel terdapat irisan responden. Selanjutnya untuk menentukan hasil estimasi pada masing-masing responden, dihitung rata-rata (average) dari hasil estimasi tiap-tiap replikasi.

Berdasarkan hasil pemilihan sampel secara acak (random) sebanyak 4 kali yang masing-masing berjumlah 300 responden, diperoleh total responden yang dapat diestimasi kemampuannya (θ) sebanyak 828 responden dari 1200 responden yang tersedia.

Dalam pemilihan secara acak sebanyak 4 kali tersebut, tidak setiap responden selalu terpilih ke dalam subkelompok sampel, sehingga hasil estimasinya tidak selalu muncul pada setiap replikasi. Dengan demikian banyaknya responden yang terestimasi hanya 828 responden dari total replikasi estimasi yang dilakukan.

Tabel 1 merupakan cuplikan hasil estimasi parameter θ (*ability estimate*) untuk responden yang terpilih pada 4 kali replikasi dengan menggunakan program *BILOG MG versi 3*.

Tabel 1. Hasil Estimasi θ Metode Bayes 4 kali Replikasi

No Uru t	Responden	Ability estimate	Ability estimate	Ability estimate	Ability estimate	Average
		r-1	r-2	r-3	r-4	
1	R0825	2.876497				2.87650
2	R1078	2.876497	2.636558			2.75653
3	R0200				2.603503	2.60350
4	R1084	2.485631				2.48563
5	R0824		2.439369			2.43937
6	R0336		2.275096		2.421203	2.34815
7	R0463		2.275096			2.27510
8	R0823				2.264263	2.26426
9	R0225			2.15952		2.15952
10	R0426		2.124017			2.12402
11	R0291				2.116184	2.11618
12	R0820				2.116184	2.11618
13	R0822	2.011487				2.01149
14	R0247			2.001583	1.956137	1.97886
15	R0315		1.95845			1.95845

.....

810	R0107		-1.686709		-1.693153	-1.68993
811	R0736		-1.686709		-1.693153	-1.68993
812	R0658				-1.693153	-1.69315
813	R1181				-1.693153	-1.69315

814	R1126	-1.758434	-1.686709			-1.72257
815	R0573	-1.758434			-1.693153	-1.72579
816	R0971	-1.758434			-1.693153	-1.72579
817	R1057	-1.758434			-1.693153	-1.72579
818	R1132	-1.758434			-1.693153	-1.72579
819	R0104	-1.758434	-1.686709	-1.735199		-1.72678
820	R0125	-1.758434	-1.686709	-1.735199		-1.72678
821	R0394			-1.735199		-1.73520
822	R0128		-1.885037	-1.930726		-1.90788
823	R1006	-1.956248	-1.885037			-1.92064
824	R0703	-1.956248			-1.890213	-1.92323
825	R0904			-1.930726		-1.93073
826	R0940			-1.930726		-1.93073
827	R0639	-2.155391	-2.084752			-2.12007
828	R0646	-2.155391		-2.128937		-2.14216

Kolom pertama pada Tabel 1 menunjukkan identitas responden, kolom kedua sampai dengan kolom keenam menunjukkan hasil estimasi kemampuan responden (*ability estimate*) pada replikasi 1 sampai dengan replikasi 5, kolom ketujuh menunjukkan rerata estimasi kemampuan. Pada Tabel 1 dapat dilihat tidak setiap responden terpilih pada tiap replikasi. Angka 0 pada kolom *ability estimate* menunjukkan bahwa responden tersebut tidak terpilih (secara random) dalam replikasi sehingga tidak diperoleh hasil estimasi kemampuannya. Estimasi akhir kemampuan responden dapat dilihat pada kolom ketujuh (*average*).

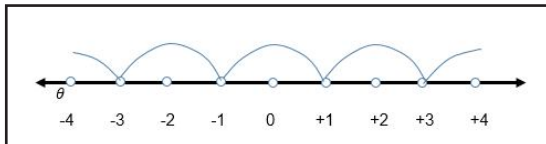
Hasil estimasi kemampuan responden pada Tabel 1 telah diurutkan dari nilai tertinggi ke nilai terendah. Tingkat kemampuan responden dapat dilihat berdasarkan nilai estimasinya yang diukur dengan satuan logit (*logarithm odd unit*). Estimasi yang bernilai positif menunjukkan tingkat kemampuan di atas rerata, sedangkan

estimasi yang bernilai negatif menunjukkan tingkat kemampuan di bawah rerata. Jadi, semakin tinggi nilai estimasi berarti semakin tinggi kemampuannya, sebaliknya semakin rendah nilai estimasi berarti semakin rendah kemampuannya. Pada Tabel 1, estimasi kemampuan tertinggi ditunjukkan oleh responden R0825 yang memiliki hasil estimasi sebesar 2.87650 logit. Hasil estimasi ini menunjukkan kemampuan responden yang tinggi karena berada hampir tiga digit di atas kemampuan rata-rata (0.0000 logit), sedangkan estimasi kemampuan terendah ditunjukkan oleh responden R0646 yang memiliki estimasi sebesar -2.14216 logit yang berarti tingkat kemampuan responden berada cukup jauh di bawah kemampuan rata-rata.

Pengukuran kemampuan siswa yang dilakukan melalui pendekatan IRT merupakan cara peskoran yang ideal. Kemampuan siswa diukur dengan mempertimbangkan taraf sukar butir soal (b). Secara teoritis, butir soal yang sulit kemungkinan akan dapat dijawab

dengan benar oleh siswa yang kemampuannya tinggi. Dengan demikian, peluang siswa yang menjawab dengan benar butir soal yang sulit dapat dianggap memiliki kemampuan yang tinggi. Menurut Naga (1992: 207) secara teoritis parameter kemampuan (θ) memiliki rentang nilai dari $-\infty$ sampai $+\infty$. Parameter kemampuan (θ) dapat memiliki nilai negatif, nol, ataupun positif. Sekalipun demikian, ada kalanya parameter kemampuan menggunakan skala nilai baku ($\mu_\theta = 0$ dan $\sigma_\theta = 1$), sehingga menurut Naga (1992: 207) secara praktis nilai baku yang masih cukup berarti untuk diamati membentang dari nilai sekitar -4,0 sampai ke sekitar +4,0.

Selanjutnya, bentangan nilai baku (antara -4,0 dan +4,0) apabila dibuat menjadi skala ordinal maka diperoleh tingkatan kemampuan siswa (*leveling*). Berdasarkan pada skala nilai baku tersebut, tingkatan kemampuan siswa dapat dikategorikan menjadi beberapa level seperti terdapat pada Grafik 3.



Grafik 3. Interval level kemampuan

Berdasarkan interval pada Grafik 3, tingkatan kemampuan siswa dapat dikategorikan menjadi 5 level kategori seperti pada Tabel 2.

Tabel 2 Level Kategori Kemampuan Siswa

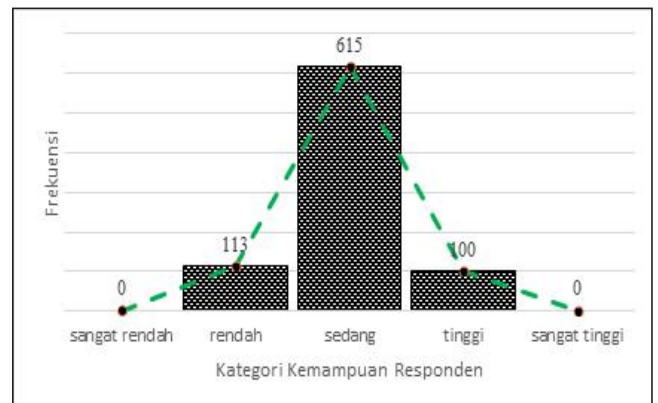
Rentang	Level Kategori
$\theta \leq -3,0$	Sangat rendah
$-3,0 < \theta \leq -1,0$	Rendah
$-1,0 < \theta \leq 1,0$	Sedang
$1,0 < \theta \leq 3,0$	Tinggi
$\theta \geq 3,0$	Sangat Tinggi

Selanjutnya, hasil estimasi kemampuan siswa secara keseluruhan ditunjukkan dalam Tabel 3.

Tabel 3. Distribusi Frekuensi Hasil Estimasi θ Metode Bayes

Rentang	Kategori	Frekuensi	Persentase
-	Sangat Rendah	0	0.00%
-2,14216 s.d -1,06891	Rendah	113	13.65%
-0,94636 s.d 0,99631	Sedang	615	74.28%
1,08084 s.d 2,87650	Tinggi	100	12.08%
-	Sangat Tinggi	0	0.00%
Total		828	100.00%

Hasil estimasi kemampuan responden sebagaimana tercantum pada Tabel 3 membentuk distribusi sebagaimana ditunjukkan pada Grafik 4.



Grafik 4. Distribusi Estimasi Kemampuan Siswa

Pada Grafik 4 tampak diagram membentuk sebuah kurva yang mendekati kurva normal. Diagram batang paling kiri menunjukkan estimasi kemampuan siswa yang paling rendah, sedangkan diagram batang paling kanan menunjukkan hasil estimasi kemampuan siswa yang paling tinggi. Berdasarkan kurva pada Grafik 4 dapat diketahui bahwa sebagian besar kemampuan responden berada pada kemampuan rerata (sedang).

Berdasarkan Tabel 3 diperoleh hasil bahwa tidak terdapat kemampuan siswa sangat ekstrim, artinya tidak ada siswa yang kemampuannya sangat rendah atau kemampuannya sangat tinggi. Sebagian besar kemampuan siswa berada pada level sedang (74,28%), siswa yang memiliki kemampuan rendah jumlahnya relatif sedikit (13,65%), demikian pula siswa dengan kemampuan tinggi jumlahnya masih relatif sedikit (12,08%).

Hasil pengukuran kemampuan siswa sebagaimana terdapat pada Tabel 3 selanjutnya dapat bermanfaat untuk digunakan sebagai bahan pertimbangan pengambilan keputusan dan kebijakan dalam pendidikan dan pembelajaran. Dengan melihat masih terdapatnya siswa yang berkemampuan rendah sebanyak 13,65%, hal ini dapat dijadikan dasar untuk memperbaiki mutu pembelajaran. Dengan demikian, kemampuan siswa dapat ditingkatnya secara signifikan.

Hasil estimasi kemampuan responden sebagaimana diuraikan di atas merupakan hasil pengukuran kemampuan siswa SMA program IPA dalam UN mata pelajaran Matematika. Hasil tersebut adalah hasil pengukuran kemampuan dengan mempertimbangkan taraf sukar butir (b) yang dijawab oleh siswa. Cara penskoran seperti ini dipandang lebih objektif karena kemampuan siswa diukur dengan mempertimbangkan bobot soal-soal yang dijawabnya. Siswa yang memiliki peluang tinggi menjawab dengan benar soal yang memiliki taraf sukar (b) tinggi, menurut IRT memiliki kemampuan yang tinggi. Sebaliknya, siswa yang memiliki kemampuan rendah, memiliki peluang yang kecil untuk dapat menjawab dengan benar

soal yang sama. Penskoran ujian dengan cara ini akan lebih objektif untuk diterapkan dalam pembelajaran.

Cara penskoran klasik yang selama ini dilakukan oleh sebagian besar guru dan pendidik merupakan cara penskoran yang hanya mempertimbangkan skor mentah (*raw score*) yang dicapai oleh siswa. Siswa yang memiliki skor tinggi dianggap memiliki kemampuan yang tinggi, sebaliknya siswa yang memiliki skor rendah dianggap memiliki kemampuan yang rendah. Padahal kondisi tersebut belum tentu demikian kenyataannya. Siswa dengan skor tinggi bisa jadi karena hanya menjawab soal-soal yang mudah sehingga jawabannya benar. Belum tentu siswa tersebut mendapatkan skor yang tinggi jika menjawab soal-soal yang sulit. Pengukuran kemampuan siswa dengan cara klasik seperti ini dianggap kurang obyektif dan menjadi tidak relevan.

Hasil penskoran cara klasik sebagaimana diuraikan di atas menunjukkan salah satu kelemahan yang terjadi dalam pengukuran estimasi kemampuan siswa. Selain kurang obyektif dalam pengukuran, penggunaan teori klasik juga dianggap memiliki reliabilitas yang rendah dalam pengukuran. Sumintono (2013: 49) menyebutkan bahwa dalam disiplin ilmu pengukuran seperti bidang pendidikan, efektifitas teori klasik mempunyai banyak keterbatasan. Ketika dua jenis tes berbeda diberikan kepada dua kelompok yang berbeda, hasil yang didapat oleh individu dan dilakukan pengujian butir soalnya tidak bisa dibandingkan.

SIMPULAN DAN SARAN

Simpulan

Hasil penelitian ini menunjukkan bahwa

74,28% kemampuan siswa SMA program IPA dalam UN mata pelajaran Matematika tahun 2015/2016 berada pada kemampuan rerata (sedang). Artinya sebagian besar siswa SMA Program IPA memiliki kemampuan yang cukup baik untuk menjawab soal-soal mata pelajaran matematika dalam UN tahun 2015/2016. Tidak ada siswa yang kemampuannya ekstrim artinya tidak ada siswa yang kemampuannya sangat rendah dan tidak ada siswa yang kemampuannya sangat tinggi.

Berdasarkan pada hasil penelitian diketahui bahwa pendekatan IRT yang menggunakan Metode Bayes merupakan cara penskoran yang ideal dan cukup akurat digunakan untuk mengestimasi kemampuan siswa. Pengukuran kemampuan siswa yang dilakukan dengan Metode Bayes selalu mempertimbangkan taraf sukar butir soal (b) yang direspon oleh siswa. Dengan demikian hasil pengukurannya akan lebih objektif dan mendekati hasil pengukuran yang sebenarnya.

Saran

Cara penskoran klasik yang selama ini masih dilakukan dengan hanya mempertimbangan skor mentah (*raw score*) yang dicapai oleh siswa dianggap kurang objektif dan tidak relevan. Sebaliknya, penskoran dengan pendekatan IRT dianggap lebih objektif. Dengan demikian, para pendidik sudah saatnya mulai menggunakan pendekatan IRT dalam melakukan penilaian di sekolah, sehingga akan diperoleh informasi yang objektif tentang perkembangan peserta didiknya dalam rangka memperbaiki mutu pembelajaran.

Berdasarkan hasil estimasi kemampuan, masih terdapat siswa SMA IPA yang berkemampuan rendah dalam mata pelajaran matematika. Informasi ini dapat dijadikan salah satu dasar pertimbangan untuk memperbaiki mutu pembelajaran khususnya mata pelajaran matematika. Dengan demikian kemampuan siswa SMA dalam mata pelajaran matematika dapat ditingkatkan secara signifikan di masa yang akan datang.

PUSTAKA ACUAN

- Azwar, Saifuddin. 2015. *Dasar-dasar Psikometrika Edisi II*. Yogyakarta: Pustaka Pelajar.
- de Ayala, R. J. 2009. *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.
- Diana, Evi Noor., dan Soehardjoepri. 2016. *Pendekatan Metode Bayesian untuk Kajian Estimasi Parameter Distribusi Log-Normal untuk Non-Informatif Prior*. Surabaya: Jurnal Sains dan Seni ITS Vol. 5 No. 2 (2016) 2337-3520 (2301-928X Print).
- du Toit, Mathilda. 2003. *IRT from SSI: BILOG MG, MULTILOG, PARSCALE, TESTFACT*. Lincolwood: Scientific Software International Inc.
- Embretson, Susan E. dan Steven P. Reise. 2000. *Item Response Theory for Psychologists*. London: Lawrence Erlbaum Associates Publishers.
- Hambleton, Ronald K., H. Swaminathan, dan H. Jane Rogers. 1991. *Fundamentals of Item Response Theory*. California: SAGE Publications Inc.
- Hartanto, Setyo., 2015. *Ujian Nasional (UN), Masih Perlukah?*. Jakarta: Jurnal [https://lppks.kemdikbud.go.id/file/Ujian_Nasional_\(UN\)Masih_Perluah.Pdf](https://lppks.kemdikbud.go.id/file/Ujian_Nasional_(UN)Masih_Perluah.Pdf).

- Hikamudin, Eviana. 2016 “*Pengaruh Metode Estimasi Terhadap Pendeteksian Kecocokan Parameter Data Lapangan dengan Model Karakteristik Logistik 1 P a r a m e t e r .*” Jakarta: Universitas Negeri Jakarta.
- Naga, Dali S. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Reskianti, Kiki., Nurtiti Sunusi, dan Nasrah Sirajang. 2014. *Estimasi Parameter Bayesian Pada Analisis Data Ketahanan Hidup Berdistribusi Eksponensial Melalui Pendekatan Self.Studi Kasus : Analisis Ketahanan Hidup Flourophores*. Makassar: Jurnal <http://repository.unhas.ac.id/handle/123456789/10274>.
- Retnawati, Heri. 2015. *Perbandingan Estimasi Kemampuan Laten Antara Metode Maksimum Likelihood dan Metode Bayes*. Yogyakarta: Jurnal Penelitian dan Evaluasi Pendidikan Volume 19, No 2, Desember 2015 (145-155).
- Setiawan, Adi. 2016. Studi Simulasi dalam Estimasi Bayesian Obyektif. Salatiga: Jurnal <http://ris.uksw.edu/download/jurnal/kode/J00287>.
- Sevilla, Consuelo G., Jesus A. Ochave, Twila G. Punsalan, Bella P. Regala, Gabriel B. Uriarte. 1993. *Pengantar Metode Penelitian*. Jakarta; Penerbit Universitas Indonesia.
- Sulistyo, Gunadi H. 2007. *Ujian Nasional (UN): Harapan, Tantangan, dan Peluang*. Malang: Jurnal Wacana Vol. 9 No. 1, April 2007 (79—106).
- Sumintono, Bambang., dan Wahyu Widhiarso. 2013. *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: Trim Komunikata Publishing House.
- Suwarto. 2005. *Pengaruh Ukuran Sampel dan Model Logistik terhadap Estimasi Parameter Item*. Sukoharjo: Jurnal Penelitian Universitas Veteran Bangun Nusantara.
- Yendra, Rado., dan Elsa Tria Noviadi. 2015. *Perbandingan Estimasi Parameter pada Distribusi Eksponensial dengan Menggunakan Metode Maksimum Likelihood dan Metode Bayesian*. Pekanbaru: Jurnal Sains Matematika dan Statistika, Vol. 1, No. 2, Juli 2015.