

## **KOMPARASI BEBERAPA METODE ESTIMASI KESALAHAN PENGUKURAN**

*Catharina Sri Wahyu Widayati*  
LPMP DIY  
cathrin\_wd@yahoo.com

### **Abstrak**

Penelitian ini bertujuan untuk mendeskripsikan besarnya estimasi kesalahan pengukuran perangkat soal Ujicoba Ujian Nasional mata pelajaran Biologi SMA. Analisis data dalam penelitian ini didasarkan pada respons siswa peserta tes ujicoba di kabupaten Bantul tahun pelajaran 2007/2008. Sumber data berupa lembar jawaban siswa yang diambil dengan teknik *stratified proportional random sampling* sebanyak 842 respons siswa dari 9 SMA negeri. Metode estimasi kesalahan pengukuran yang digunakan dalam penelitian ini adalah 1) metode Thorndike, 2) metode Polynomial, 3) pendekatan Binomial Lord dengan modifikasi Keats, 4) pendekatan compound Binomial, 5) metode analisis varians, dan 6) metode teori respons butir. Hasil analisis data menunjukkan bahwa besarnya estimasi kesalahan pengukuran perangkat soal tes ujicoba ujian nasional mata pelajaran biologi SMA tahun pelajaran 2007/2008 dengan menggunakan metode Thorndike, Polynomial, Binomial dengan modifikasi Keats, Compound Binomial, Analisis Varians, dan teori respons butir berturut-turut sebagai berikut: 2,96, 2,99, 2,87, 2,87, 2,83, dan 2,81, sedangkan hasil perhitungan *root mean square error* (RMSE) untuk masing-masing metode berturut-turut sebagai berikut: 0,02869, 0,02904, 0,02891, 0,02873, 0,02854, dan 0,02848. Harga RMSE terkecil diperoleh dengan metode teori respons butir, sehingga metode ini merupakan metode yang paling tepat.

Kata kunci: *metode estimasi kesalahan pengukuran*

## **COMPARISON OF SEVERAL METHODS OF THE MEASUREMENT ERROR ESTIMATION**

*Catharina Sri Wahyu Widayati*  
LPMP DIY  
cathrin\_wd@yahoo.com

### **Abstract**

This study aims to find out 1) the magnitudes of the measurement error estimation of the test for the national examination tryout of Biology for senior high schools in the Academic Year of 2007/2008 in Bantul Regency by employing six methods and 2) the most appropriate method of the measurement error estimation of the test. Data were analyzed based on the students' responses. The data source was 842 students' answer sheets selected by using the stratified proportional random sampling technique from 9 senior high schools in Bantul Regency. The methods of the measurement error estimation employed in this study included the Thorndike method, Polynomial method, Lord's binomial method with Keats' modification, compound Binomial method, method of analysis of variance, and method of item response theory. The results showed that the magnitudes of the measurement error estimation of the test are 2.96, 2.99, 2.87, 2.87, 2.83, and 2.81 respectively for the methods of Thorndike, Polynomial, Lord's binomial with Keats' modification, compound Binomial, analysis of variance, and item response theory. The values of root mean square error (RMSE) for each method are 0.02869, 0.02904, 0.02891, 0.02873, 0.02854, and 0.02848. The smallest RMSE is obtained by using the method of item response theory. Therefore, it can be concluded that is the most appropriate method of the measurement error estimation of the test.

Key words: *methods of the measurement error estimation*

## **Pendahuluan**

Pengukuran mempunyai peranan penting dalam bidang pendidikan, terutama dalam proses pembelajaran. Dengan melakukan pengukuran, guru akan mengetahui tingkat kemampuan siswa dalam bidang tertentu. Menurut Allen & Yen (1979: 2) pengukuran didefinisikan sebagai penetapan suatu angka terhadap suatu subjek dengan cara yang sistematis. Hasil pengukuran yang berupa angka/skor diharapkan mencerminkan kemampuan peserta tes yang sebenarnya. Namun demikian ada kemungkinan terdapat perbedaan antara skor hasil pengukuran dengan skor sebenarnya. Ada dua macam kemungkinan, pertama, mungkin skor hasil pengukuran lebih rendah daripada skor yang sebenarnya. Kedua, mungkin skor yang diberikan kepada peserta tes lebih tinggi daripada skor sebenarnya. Bila kemungkinan itu terjadi, maka berarti terjadi kesalahan pengukuran.

Menurut Azwar (2007: 5) kesalahan pengukuran menunjuk pada sejauh mana inkonsistensi hasil pengukuran terjadi apabila pengukuran dilakukan ulang pada kelompok subjek yang sama. Semakin kecil harga kesalahan pengukuran maka pengukuran semakin cermat dan semakin dapat dipercaya. Ada dua macam kesalahan dalam pengukuran yaitu kesalahan sistematis dan kesalahan acak.

Kesalahan yang bersifat sistematis disebabkan oleh orang yang mengukur atau alat ukur/instrumennya (Djemari Mardapi, 2000: 5). Kesalahan pengukuran yang sistematis adalah kesalahan yang secara konsisten mempengaruhi hasil pengukuran. Kesalahan pengukuran sistematis ini disebabkan karena soal tes yang terlalu mudah atau terlalu sukar, selain itu ada guru yang pemurah dan ada yang mahal dalam memberi skor. Kesalahan sistematis tidak bisa diestimasi besarnya. Sedangkan kesalahan yang bersifat acak tidak memiliki pola secara sistematis. Kesalahan acak disebabkan antara lain karena kesalahan dalam menentukan sampel isi tes, dan adanya variasi emosi seseorang yang bersifat acak. Besarnya kesalahan pengukuran acak ini yang dapat diestimasi.

Ada beberapa sumber kesalahan yang mempengaruhi hasil pengukuran yang meliputi alat ukur, objek yang diukur, lingkungan pengukuran, dan subjek yang mengukur (Djemari Mardapi, 1999: 4). Dalam bidang pendidikan sumber kesalahan pengukuran lebih banyak dan lebih kompleks, karena baik objek yang diukur maupun subjek yang mengukur adalah manusia. Menurut Feldt & Brennan (1989: 107) pengukuran dalam bidang pendidikan bersifat tidak langsung, hasilnya ditentukan oleh kondisi fisik dan psikologis yang diukur dan yang mengukur. Hal ini disebabkan karena kondisi fisik dan emosi seseorang selalu bervariasi dari waktu ke waktu.

Sumber-sumber penyebab terjadinya kesalahan pengukuran sangat sulit untuk dikendalikan, tetapi dapat diusahakan agar kesalahan pengukuran dapat diminimalkan, sehingga perolehan skor dapat mencerminkan kemampuan peserta tes yang sebenarnya (Djemari Mardapi, 1999: 15). Di antara sumber-sumber kesalahan pengukuran itu, nampaknya yang paling mudah dikontrol adalah faktor alat yang dipergunakan untuk mengukur. Oleh karena itu, dalam usaha memperkecil kesalahan-kesalahan pengukuran, diperlukan alat ukur yang baik.

Alat ukur yang baik adalah alat ukur yang memiliki bukti kesahihan dan keandalan. Bukti kesahihan alat ukur dilihat pada kesesuaian antara definisi operasional dari konsep yang akan diukur dengan materi pertanyaan pada alat ukur. Bukti kesahihan alat ukur meliputi kesahihan isi, konstruk, dan kriteria.

Bukti keandalan suatu alat ukur dapat dilihat pada besarnya indeks keandalan. Besarnya indeks keandalan digunakan untuk menghitung besarnya kesalahan pengukuran. Semakin tinggi indeks keandalan semakin kecil kesalahan pengukuran, demikian juga sebaliknya. Dengan mengetahui besarnya kesalahan pengukuran maka akan dapat ditafsirkan apakah perbedaan skor yang ada merupakan perbedaan yang nyata atau hanya suatu kebetulan.

Ada dua teori yang saat ini banyak digunakan dalam pengukuran, yaitu teori tes klasik dan teori respons butir. Kedua teori ini digunakan untuk menaksir kemampuan seseorang dengan tingkat kesalahan sekecil mungkin. Asumsi yang digunakan oleh kedua teori ini berbeda sehingga

menghasilkan teknik estimasi yang berbeda.

Estimasi kesalahan pengukuran pada teori tes klasik menggunakan reliabilitas tes. Semakin besar reliabilitasnya akan semakin kecil kesalahan pengukurannya. Reliabilitas tes dalam teori repons butir dinyatakan dengan menggunakan fungsi informasi tes, sehingga untuk menentukan besarnya kesalahan pengukuran menggunakan fungsi informasi tes. Semakin tinggi fungsi informasi tes, kesalahan pengukuran semakin rendah.

Besarnya kesalahan pengukuran dapat diestimasi dengan beberapa metode. Ada beberapa metode estimasi yang dikembangkan yaitu metode Thorndike, metode Polynomial, metode Binomial Lord dengan modifikasi Keats, metode compound Binomial, metode analisis varians, dan metode teori respons butir (Feldt & Brennan, 1989: 123-124).

### 1) Metode Thorndike

Thorndike pengukuran untuk tes belah dua. Metode ini pada prinsipnya membagi tes menjadi dua bagian, dan total skor yang diperoleh seseorang merupakan penjumlahan skor dari dua tes paralel, masing-masing setengah dari tes keseluruhan.

Dengan mengikuti model teori tes klasik, masing-masing belahan tes dan total skor tes dianggap sebagai penjumlahan komponen “*true*” dan komponen “*error*”:

$$\begin{aligned} X_1 &= T_1 + E_1 \\ X_2 &= T_2 + E_2 \\ X &= X_1 + X_2 \\ &= (T_1 + T_2) + (E_1 + E_2) = T + E \end{aligned} \tag{1}$$

Konsisten dengan teori tes klasik, korelasi antara  $E_1$  dan  $E_2$  diasumsikan sama dengan nol.

$$\sigma = (\sigma_{E_1}^2 + \sigma_{E_2}^2)^{1/2} \tag{2}$$

Selisih antara skor belahan tes, yaitu  $X_1 - X_2$ , sama dengan  $(T_1 - T_2) + (E_1 - E_2)$ . Untuk belahan tes paralel,  $T_1 = T_2$  atau  $T_1 - T_2 = 0$ , sehingga

$$\sigma_{(X_1 - X_2)} = \sigma_{(E_1 - E_2)}$$

$$\begin{aligned} &= (\sigma_{E_1}^2 + \sigma_{E_2}^2)^{1/2} \\ &= \sigma_E \end{aligned} \tag{3}$$

Oleh karena itu, kesalahan baku total dapat diestimasi, yaitu berupa simpangan baku beda skor dua tes yang paralel tersebut.

## 2) Metode Polinomial

Pada dasarnya metode Polinomial berasal dari metode Thorndike yang dikembangkan oleh Mollenkopf (1949). Mollenkopf menggunakan teknik regresi untuk mencari besarnya varians kesalahan. Mollenkopf mengusulkan bahwa kuadrat selisih skor belahan tes untuk masing-masing individu dianggap sebagai nilai yang diprediksikan oleh teknik regresi dari total skor. Jadi  $Y = (X_1 - X_2)^2$  dianggap sebagai *criterion* yang diprediksikan menggunakan pendekatan regresi kuadrat terkecil.

Persamaan yang digunakan pada metode polinomial:

$$\hat{Y} = a_0 + a_1(X) + a_2(X^2) + \dots + a_k(X^k) \tag{4}$$

Dengan  $\hat{Y}$  diinterpretasikan sebagai estimasi  $\mu_y$  untuk populasi individu yang memperoleh nilai X tertentu. Namun demikian, rata-rata nilai Y adalah rata-rata kuadrat selisih skor belahan tes. Dengan belahan tes paralel, rata-rata kuadrat selisih adalah varians dari selisih tersebut. Nilai ini, seperti yang ditunjukkan Thorndike, adalah varians kesalahan pengukuran untuk tes utuh. Oleh karena itu,  $(\hat{Y})^{1/2}$  untuk nilai X tertentu adalah estimasi kesalahan pengukuran pada skor X.

## 3) Metode Binomial Lord dengan modifikasi Keats

Metode estimasi ini didasarkan pada model kesalahan binomial. Individu tertentu  $i$  dianggap mampu menjawab proporsi tertentu  $i$  dari populasi item keseluruhan. Dasar utama pendekatan ini adalah bahwa kesalahan baku pengukuran merupakan simpangan baku skor peserta tes

yang mengikuti tes paralel yang banyak. Lord mengusulkan penggunaan proporsi jawaban benar sebagai estimasi  $\hat{z}$ . Dia juga merekomendasikan koreksi untuk bias yang diketahui dalam varians dari sampel tertentu.

$$S_E = \left[ \frac{X(k-X)}{k-1} \right]^{1/2} \tag{5}$$

Persamaan 5 merupakan estimasi kesalahan pengukuran pada level skor X untuk tes total.

Namun demikian, persamaan (5) memiliki kelemahan. Persamaan ini tidak mempertimbangkan kecocokan form tes dalam hal isi, tingkat kesulitan, dan karakteristik lain. Oleh karena itu persamaan (3) nampaknya *overestimate* terhadap kesalahan pengukuran pada nilai skor X. Kenyataan bahwa rata-rata kuadrat nilai persamaan (5) mengarah pada KR-21 sebagai koefisien reliabilitas tes menguatkan keyakinan bahwa persamaan tersebut *overestimate* terhadap kesalahan pengukuran.

Keats (1957) mengusulkan modifikasi persamaan (5) dengan mengalikan sisi kanan persamaan tersebut dengan suatu konstanta, yang akan menghasilkan nilai rata-rata yang konsisten dengan koefisien reliabilitas yang lebih tepat. Hasil modifikasi keats tersebut disajikan:

$$S_E = \left[ \left( \frac{X(k-X)}{k-1} \right) \left( \frac{1-r_{xx'}}{1-r_{21}} \right) \right]^{1/2} \tag{6}$$

Dengan  $r_{xx'}$  adalah estimasi reliabilitas tes, dan  $r_{21}$  adalah KR-21 untuk tes. Keats merekomendasikan koefisien bentuk paralel untuk  $r_{xx'}$ , tetapi dalam praktiknya mungkin perlu menggunakan koefisien belah dua atau Cronbach's (1951) alfa.

#### 4) Metode Compound Binomial

Metode ini beranggapan bahwa tes yang paralel pada dasarnya adalah suatu proses pemilihan sampel berstrata dari butir soal. Lord dan Feldt mengembangkan metode ini dengan menggunakan teori statistika untuk memperoleh formula persamaan kesalahan baku berdasarkan frekuensi

yang diperoleh dari sampel berstrata. Kesalahan baku pengukuran model ini disajikan dengan persamaan:

$$S_{E(i)} = \left[ \sum_{h=1}^c \frac{X_{ih} (k_h - X_{ih})}{k_h - 1} \right]^{1/2} \quad (7)$$

Keterangan:

$S_{E(i)}$  = kesalahan pengukuran untuk orang ke-i

$X_{ih}$  = skor orang ke-i pada kelompok item yang berhubungan dengan kategori h dari spesifikasi tes

c = jumlah kategori item

$k_h$  = jumlah item dalam kategori h

Untuk menggunakan persamaan ini, tes harus diskor untuk tiap kategori item seolah-olah kategori adalah subtes dari instrumen keseluruhan. Aplikasi persamaan ini menghasilkan estimasi kesalahan pengukuran masing-masing peserta tes.

#### 5) Metode Analisis Varians

Hoyt (1941) mengembangkan suatu metode dengan menggunakan teori komponen varians. Matriks skor item tes dianalisis untuk memperoleh rata-rata kuadrat peserta tes ( $MS_s$ ), rata-rata kuadrat item ( $MS_I$ ), dan rata-rata kuadrat interaksinya ( $MS_{SI}$ ), dan reliabilitas tes dapat diestimasi dari rata-rata kuadrat tersebut. Dalam hal tertentu, varians *error* untuk tes dengan k butir soal dapat diperkirakan oleh  $k(MS_{SI})$ . Apabila individu dikelompokkan berdasarkan interval skor total, kesalahan baku pengukuran untuk skor dapat diestimasi dengan persamaan:

$$S_E = \left[ k(MS_{SI}) \right]^{1/2} \quad (8)$$

Estimasi ini terkait erat dengan estimasi Thorndike (1951), dan memiliki kelemahan berupa ketidakstabilan yang muncul dari jumlah peserta tes yang kecil pada interval ekstrim.

6) Metode Teori respons butir

Metode ini dirumuskan berdasarkan teori kurva respons butir yang dikembangkan oleh Lord. Varians kesalahan pengukuran seseorang diperoleh melalui estimasi kemampuan peserta tes ( $\theta_j$ ), dan evaluasi fungsi varians kesalahan baku pengukuran untuk individu. Metode ini menuntut ukuran sampel yang besar agar hasilnya akurat, tetapi dapat mengestimasi kesalahan baku pengukuran untuk setiap individu.

$$S_{E(i)} = \left\{ \sum_j^k [P_j(\theta_i)][1 - P_j(\theta_i)] \right\}^{1/2} \tag{9}$$

Dengan  $P_j(\theta_j)$  adalah nilai fungsi untuk item  $j$  pada level kemampuan  $\theta_j$ , dari subjek  $i$ . Seperti pada pendekatan Binomial Compound, metode ini menghasilkan estimasi untuk setiap peserta tes. Untuk menggunakan metode ini, peserta tes dikelompokkan menurut level skor total mereka, nilai rata-rata  $S_{E(i)}$  dihitung untuk peserta tes pada setiap interval, dan akar kuadrat rata-rata ditentukan untuk interval korespondensi satu persatu antara nilai  $X$  dan nilai  $\theta$ . Setiap pasangan spesifik ( $X_0, \theta_0$ ) diperoleh dengan penentuan nilai  $\theta_0$ , yang memenuhi hubungan:

$$X_0 = \sum_j^k P_j(\theta_0) \tag{10}$$

Kuadrat kesalahan pengukuran pada nilai  $\theta_0$ , seperti dihitung dengan persamaan (24), kemudian dihubungkan dengan nilai skor mentah  $X_0$ .

Setiap metode estimasi mempunyai asumsi dan formulasi yang berbeda, sehingga kemungkinan akan menghasilkan estimasi yang berbeda pula. Hasil estimasi tersebut tergantung pada karakteristik data. Untuk mengetahui metode menerapkannya pada suatu perangkat soal. Dalam hal ini perangkat soal yang digunakan adalah soal Ujicoba Ujian Nasional mata pelajaran Biologi SMA Tahun Pelajaran 2007/2008 di Kabupaten Bantul.

Metode yang paling tepat untuk mengestimasi besarnya kesalahan pengukuran didasarkan pada *root mean square error* (RMSE) yang dihasilkan

oleh masing-masing metode. RMSE digunakan untuk membandingkan metode-metode estimasi yang digunakan, yaitu untuk menentukan metode estimasi yang paling akurat. Menurut Kolen & Brennan (1995: 231) RMSE dapat ditentukan dengan persamaan berikut:

$$RMSE \cong \sqrt{\frac{\sigma_Y^2 [1 - \rho_{X,Y}]}{N} \left\{ 2 + [1 + \rho_{X,Y}] \left[ \frac{X_i - \mu_X}{\sigma_X} \right]^2 \right\}} \quad (11)$$

Keterangan:

RMSE = *root mean square error*

$N$  = banyaknya data

$\rho_{X,Y}$  = korelasi antara  $X$  dan  $Y$

$\sigma_Y^2$  = varians skor  $Y$

$\sigma_X$  = simpangan baku skor  $X$

$\mu_X$  = rata-rata skor  $X$

Keakuratan metode estimasi kesalahan pengukuran diindikasikan dengan adanya RMSE yang kecil. Metode estimasi yang mempunyai RMSE lebih kecil dikatakan lebih akurat daripada metode estimasi yang mempunyai RMSE lebih besar.

### Metode Penelitian

Subjek dalam penelitian ini adalah lembar jawaban siswa kelas XII SMA Negeri yang telah mengikuti Ujicoba Ujian Nasional mata pelajaran Biologi Tahun Pelajaran 2007/2008 di Kabupaten Bantul untuk tipe tes objektif pilihan ganda. Banyaknya SMA Negeri di Kabupaten Bantul berjumlah 19 sekolah.

Beberapa sekolah dipilih sebagai sumber data untuk keperluan penelitian dengan teknik *stratified proportional random sampling*. Teknik ini dilakukan dengan membagi sekolah-sekolah dalam beberapa strata dan mengambil subjek yang terdapat dalam tiap strata secara proporsional dan

random. Secara lengkap SMA Negeri di Kabupaten Bantul yang diambil datanya sebagai subjek dalam penelitian ini disajikan dalam Tabel 1.

Tabel 1. Daftar SMA Kabupaten Bantul yang Menjadi Sumber Data

No	Nama Sekolah	Kategori	Jumlah Data
1	SMA Negeri 1 Bantul	Tinggi	42
2	SMA Negeri 2 Bantul		100
3	SMA Negeri 1 Kasihan		80
4	SMA Negeri 1 Pleret		62
5	SMA Negeri 3 Bantul		79
6	SMA Negeri 1 Sewon		104
7	SMA Negeri 1 Jetis		76
8	SMA Negeri 1 Sedayu		142
9	SMA Negeri 1 Banguntapan		157

Data berupa respon siswa kelas XII yang mengikuti Ujicoba Ujian Nasional mata pelajaran Biologi Tahun Pelajaran 2007/2008 sebanyak 842 respon diperoleh dari sejumlah SMA Negeri di Kabupaten Bantul yang dijadikan subjek penelitian. Data yang terkumpul berupa respon siswa terhadap perangkat soal Biologi sebanyak 40 butir soal bentuk pilihan ganda dengan 5 pilihan jawaban yang berupa data mentah.

Data penelitian ini dikumpulkan dengan memeriksa dan mencatat dokumen yang berupa lembar jawaban peserta Ujicoba Ujian Nasional mata pelajaran Biologi SMA Tahun Pelajaran 2007/2008 di Kabupaten Bantul yang telah dijadikan sumber data. Dokumen tersebut diperoleh dari guru mata pelajaran Biologi Kelas XII masing-masing sekolah.

Data diestimasi dengan menggunakan metode Thorndike, metode Polynomial, metode Binomial Lord dengan modifikasi Keats, metode *compound Binomial*, metode analisis varians, dan metode teori respons butir untuk menentukan besar kesalahan pengukuran. Pelaksanaan analisis dilakukan dengan bantuan komputer.

Metode yang terbaik untuk mengestimasi kesalahan baku pengukuran perangkat soal ujicoba ujian nasional mata pelajaran Biologi SMA Tahun Pelajaran 2007/2008 di Kabupaten Bantul adalah metode yang

memberikan harga *root mean square error* (RMSE) yang terkecil. RMSE untuk masing-masing metode estimasi dihitung menggunakan persamaan (11).

### Hasil Penelitian dan Pembahasan

Berdasarkan hasil analisis data dengan menggunakan metode Thorndike, metode Polynomial, metode Binomial dengan modifikasi Keats, metode *Compound Binomial*, metode Analisis Varians, dan metode teori respons butir (IRT) dapat diketahui bahwa estimasi kesalahan pengukuran terkecil untuk perangkat soal ujicoba ujian nasional mata pelajaran Biologi SMA tahun pelajaran 2007/2008 diperoleh dengan metode IRT. Rangkuman hasil estimasi kesalahan pengukuran selengkapnya disajikan pada Tabel 2.

Tabel 2. Hasil Estimasi Kesalahan Pengukuran dengan Beberapa Metode

No	Metode	SEM	RMSE
1	Thorndike	2,96	0,02869
2	Polynomial	2,99	0,02904
3	Modifikasi Keats	2,87	0,02891
4	Compound Binomial	2,87	0,02873
5	Analisis Varian	2,83	0,02854
6	IRT	2,81	0,02848

Estimasi kesalahan pengukuran dengan metode Thorndike diperoleh dengan membagi tes menjadi dua belahan tes yang  $\tau$ -*equivalent*. Jika sebuah tes dapat dibagi menjadi 2 belahan yang  $\tau$ -*equivalent*, maka varians selisih antara kedua belahan sama dengan varians kesalahan untuk tes tersebut. Dalam penelitian ini tes dibelah menjadi 2 bagian sama panjang berdasarkan tingkat kesukaran butir tes, masing-masing belahan terdiri dari 20 butir soal. Harga tingkat kesukaran butir tes diperoleh dari hasil analisis data menggunakan program Iteman.

Metode Polynomial menggunakan teknik regresi untuk mencari besarnya varians kesalahan. Tes dibelah dua sama panjang, kuadrat selisih skor dua belahan untuk masing-masing peserta dianggap sebagai nilai yang diprediksikan oleh teknik regresi dari skor total. Dengan bantuan program

SPSS dapat diketahui bahwa persamaan regresi yang cocok dengan data adalah persamaan regresi kuadratik.

Model binomial merupakan versi empirik dari IRT. Sebagai pengganti penggunaan pendekatan *maximum likelihood* untuk estimasi  $\theta$  masing-masing individu, metode binomial lebih menekankan pada skor tampak untuk masing-masing peserta tes, mengelompokkan individu-individu dengan  $X_0$  yang sama. Metode ini menggunakan proporsi jawaban benar sebagai estimasi proporsi penguasaan domain. Varians kesalahan dicari dengan menghitung varians jumlah jawaban benar dengan koreksi untuk bias dalam varians.

Analisis dengan metode *Compound Binomial* menyaratkan adanya pemilihan sampel berstrata dari butir soal. Perangkat tes dikelompokkan menjadi beberapa strata berdasarkan tingkat kesukaran butir soal. Varians kesalahan untuk tes keseluruhan dapat diestimasi dengan menjumlahkan hasil estimasi dari masing-masing strata.

Metode Analisis Varians memandang distribusi item keseluruhan sebagai data pada suatu desain eksperimen faktorial. Jika bagian-bagian tes dan peserta dianggap sebagai sampel dari domain, rataan kuadrat dari interaksi merupakan estimasi terhadap varians kesalahan untuk tiap bagian. Sedangkan varians kesalahan untuk keseluruhan tes dicari dengan mengalikan rataan kuadrat interaksi dengan banyaknya bagian tes.

Analisis dengan metode teori respons butir dilakukan dengan bantuan program Bilog-MG untuk mencari harga estimasi parameter butir dan parameter kemampuan ( $\theta$ ). Perhitungan estimasi kesalahan pengukuran dilakukan dalam hubungannya dengan skor mentah. Varians kesalahan untuk skor mentah tertentu didefinisikan sebagai penjumlahan peluang binomial menjawab benar untuk setiap item pada perangkat tes dengan tingkat kemampuan yang berhubungan dengan skor mentah tersebut. Masing-masing skor mentah, memiliki hubungan dengan hanya satu estimasi kemampuan.

Estimasi kesalahan pengukuran yang diperoleh dari masing-masing metode dapat digunakan untuk mengestimasi interval skor murni. Lebar atau sempitnya interval skor murni pada taraf kepercayaan tertentu tergantung pada besar kecilnya kesalahan baku pengukuran. Kesalahan

baku pengukuran membantu menyatakan batas kesalahan yang terjadi pada setiap tes dan secara khusus berguna untuk mengestimasi skor peserta tes yang sebenarnya. Kesalahan baku pengukuran membantu menginterpretasikan kedekatan antara skor tampak dan skor sebenarnya, dalam hal ini menetapkan interval kepercayaan. Semakin kecil kesalahan baku pengukuran semakin sempit interval kepercayaan skor sebenarnya, yang berarti hasil tes semakin cermat. Sebaliknya, semakin besar kesalahan baku pengukuran maka akan semakin lebar pula interval kepercayaan yang terjadi dan hasil pengukuranpun semakin tidak cermat.

Hasil estimasi kesalahan pengukuran dengan berbagai metode (Tabel 2) menunjukkan bahwa terdapat variasi harga estimasi kesalahan pengukuran perangkat soal Ujicoba Ujian Nasional Mata Pelajaran Biologi jenjang SMA Tahun Pelajaran 2007/2008. Estimasi kesalahan pengukuran terkecil diperoleh dengan metode IRT, diikuti oleh metode Analisis Varians, metode Binomial dengan modifikasi Keats, metode *Compound Binomial*, dan metode Thorndike, sedang estimasi kesalahan pengukuran terbesar diperoleh dengan metode Polynomial.

Berdasarkan hasil perhitungan *root mean square error* (RMSE), metode IRT mempunyai nilai RMSE terkecil. Nilai RMSE terkecil merupakan kriteria metode estimasi kesalahan pengukuran yang akurat. Hal ini berarti metode IRT merupakan metode estimasi kesalahan pengukuran yang paling tepat untuk perangkat soal tersebut. Meskipun asumsi yang digunakan dalam IRT berbeda, akan tetapi dalam penelitian ini estimasi kesalahan pengukuran dengan metode IRT dilakukan dalam *terms* skor mentah sama seperti metode-metode yang lain, sehingga hasilnya dapat dibandingkan secara empirik.

Hasil analisis dengan menggunakan teori tes klasik sangat tergantung pada karakteristik peserta yang diukur. Apabila perangkat soal tersebut diujikan pada sekelompok peserta yang lain, maka hasil analisisnya akan berubah sesuai dengan karakteristik peserta tesnya. Meskipun demikian, dengan cara ini dapat memberikan rambu-rambu terhadap pembuatan soal yang baik.

Kelemahan-kelemahan yang terdapat dalam teori tes klasik dapat diatasi dengan teori respons butir. Teori respons butir melepaskan

keterikatan antara butir soal dengan peserta tes. Karakteristik peserta akan tetap sama walau mengerjakan butir soal yang berbeda, demikian pula karakteristik butir soal juga akan tetap sama walau dikerjakan oleh peserta yang berbeda kemampuannya.

Secara ringkas dapat dikatakan bahwa semua metode mempunyai dasar teoritis yang sama. Meskipun metode-metode estimasi nampak berbeda, tetapi memiliki variasi dalam konsep yang sama yaitu jika peserta tes dapat dikelompokkan berdasar skor murninya, varians kesalahannya akan sama dengan  $\sum P_j (1 - P_j)$  dimana  $P_j$  merupakan proporsi menjawab benar untuk item  $j$ . Nampak jelas bahwa kesamaan data empiris untuk beberapa metode mencerminkan kesamaan konsep diantara metode-metode tersebut. Pemilihan berbagai metode tergantung pada pertimbangan praktis dan pada kesukaan pengguna terhadap logika yang melandasi masing-masing metode.

## **Simpulan**

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan sebagai berikut:

1. Besarnya estimasi kesalahan pengukuran perangkat soal Ujicoba Ujian Nasional mata pelajaran biologi SMA tahun pelajaran 2007/2008 dengan menggunakan metode Polynomial, Thorndike, Binomial dengan modifikasi Keats, *Compound Binomial*, Analisis Varians, dan teori respons butir (IRT) berturut-turut sebesar: 2,99, 2,96, 2,87, 2,87, 2,83, dan 2,81.
2. Estimasi kesalahan pengukuran terkecil diperoleh dengan metode IRT, diikuti oleh metode Analisis Varians, metode Binomial dengan modifikasi Keats, metode *Compound Binomial*, dan metode Thorndike, sedangkan estimasi kesalahan pengukuran terbesar diperoleh dengan metode Polynomial, walaupun secara umum perbedaannya tidak terlalu jauh.
3. Metode teori respons butir memberikan hasil perhitungan *root mean square error* (RMSE) paling kecil dibandingkan dengan metode-metode estimasi yang lain, dengan demikian metode teori respons butir merupakan metode yang paling tepat untuk mengestimasi kesalahan

pengukuran perangkat soal ujicoba ujian nasional mata pelajaran biologi SMA tahun pelajaran 2007/2008.

### **Daftar Pustaka**

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company.
- Azwar, S. (2007). *Reliabilitas dan validitas* (edisi ke-3). Yogyakarta: Pustaka Pelajar.
- Djemari Mardapi. (1999). *Estimasi kesalahan pengukuran dalam bidang pendidikan dan implikasinya pada ujian nasional*. Pidato Pengukuhan Guru Besar Madya Ilmu Evaluasi Pendidikan Teknologi dan Kejuruan pada Fakultas Teknik Universitas Negeri Yogyakarta.
- \_\_\_\_\_. (2000). *Beberapa masalah evaluasi pada pendidikan biologi*. Makalah Seminar Lokakarya Pendidikan Biologi FMIPA, UNY.
- \_\_\_\_\_. (2004). *Penyusunan Tes Hasil Belajar*. Program Pascasarjana Universitas Negeri Yogyakarta.
- Feldt, L. S. & Brennan, R. L. (1989). Reliability. Dalam Linn (1989). *Educational measurement*. London: Collier Macmillan Publisher.
- Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*. 153-160
- Keats, J. A. (1957). Estimation of error variances of test scores. *Psychometrika*. 29-41
- Kolen, M. J. & Brennan, R. L. (1995). *Tes equating: methods and practices*. New York: Verlag.
- Mollenkopf, W. G. (1949). Variation of the standard error of measurement. *Psychometrika*. 189-229.
- Thorndike, R. L. (1951). Reliability. Dalam Lindquist, E. F. (1951). *Educational measurement*. Washington DC: American council on education.