

ESTIMASI KESALAHAN PENGUKURAN EBTANAS MATA PELAJARAN IPA SLTP DI KOTAMADIA SURAKARTA

Oleh:

Nurul Kamilati dan Djemari Mardapi

Abstrak

Penelitian ini bertujuan untuk mengetahui estimasi kesalahan pengukuran perangkat EBTANAS mata pelajaran IPA SLTP di Kotamadia Surakarta tahun pelajaran 1997/1998 dengan menggunakan teori tes klasik dan teori respons butir. Selain itu ingin mengetahui hubungan harga estimasi kesalahan pengukuran teori tes klasik dengan interval skor murni peserta EBTANAS dan hubungan harga estimasi kesalahan pengukuran teori respons butir dengan interval tingkat kemampuan peserta EBTANAS.

Perangkat EBTANAS mata pelajaran IPA yang dianalisis terdiri dari 45 butir soal pilihan ganda. Populasi penelitian ini berjumlah 13.681 lembar jawaban. Pengambilan sampel dilakukan dengan teknik *cluster stratified proportional systematic sampling* diperoleh sampel berjumlah 3.681 lembar jawaban. Data penelitian yang berupa respons jawaban peserta EBTANAS diperoleh melalui dokumentasi. Estimasi kesalahan pengukuran dengan teori tes klasik Model Feldt terhadap perangkat EBTANAS memperhatikan klasifikasi dan perangkat EBTANAS tanpa memperhatikan klasifikasi dilakukan dengan membagi kedua perangkat menjadi dua belahan yang tidak sama panjang, asalkan memenuhi asumsi homogenitas dari sisi tingkat kesukaran butir soal. Estimasi kesalahan pengukuran teori respons butir model logistik tiga parameter perangkat EBTANAS yang semua butir soalnya dianalisis dan perangkat EBTANAS yang butir-butirnya cocok dengan model logistik tiga parameter dilakukan dengan menghitung harga fungsi informasi butir soal pada tiap tingkat kemampuan peserta θ dari -4 sampai +4. Selanjutnya diperoleh harga fungsi informasi perangkat EBTANAS dan estimasi kesalahan pengukurannya.

Hasil penelitian menunjukkan bahwa menurut teori tes klasik Model Feldt, harga estimasi kesalahan pengukuran perangkat EBTANAS memperhatikan klasifikasi dan tanpa memperhatikan klasifikasi berturut-turut adalah 2.2791 dan 2.9233. Interval kepercayaan skor murni peserta EBTANAS pada tingkat kepercayaan $\alpha = 0,05$ adalah $X - 4,4670 \leq T \leq X + 4,4670$ untuk perangkat EBTANAS memperhatikan klasifikasi dan $X - 5,7297 \leq T \leq X + 5,7297$ untuk perangkat EBTANAS tanpa memperhatikan klasifikasi. Estimasi kesalahan pengukuran dengan teori respons butir model logistik tiga parameter menunjukkan estimasi kesalahan pengukuran terendah terjadi pada tingkat kemampuan peserta $\theta = 0,5$. Interval tingkat kemampuan peserta EBTANAS pada tingkat kepercayaan $\alpha = 0,05$ dan tingkat kemampuan $\theta = 0,5$ untuk perangkat EBTANAS yang semua butir soalnya dianalisis dan perangkat EBTANAS yang butir-butir soalnya cocok dengan model logistik tiga parameter berturut-turut adalah -0,1256 sampai dengan 1,1256 dan -0,1504 sampai dengan 1,1504.

Pendahuluan

EBTANAS merupakan salah satu jenis evaluasi prestasi belajar yang sangat penting. Tujuan pelaksanaan EBTANAS adalah untuk memperoleh informasi tentang mutu hasil pendidikan secara nasional; mengukur pencapaian hasil belajar siswa baik sekolah/madrasah negeri maupun sekolah/madrasah swasta; memperoleh umpan balik bagi sekolah/madrasah dalam penyempurnaan program dan upaya peningkatan mutu pendidikan secara berkelanjutan; memperoleh gambaran perbandingan mutu pendidikan pada sekolah/madrasah, antarsekolah/madrasah, dan antarwilayah dari tahun ke tahun; menjadi bahan penentuan kebijakan pembinaan sekolah/madrasah; sebagai bahan pertimbangan dalam memberikan Surat Tanda Tamat Belajar (STTB) dan seleksi masuk ke jenjang pendidikan yang lebih tinggi (Depdikbud; Dinas P & K; Depag: 1998).

Meskipun pelaksanaan EBTANAS telah dilakukan beberapa kali namun hasil yang diperoleh belum memuaskan. Hal ini ditunjukkan dengan perolehan Nilai Ebtanas Murni (NEM) yang rendah, khususnya pada bidang studi Ilmu Pengetahuan Alam (IPA). Berdasarkan data dari Kantor Departemen Pendidikan dan Kebudayaan Kotamadia Surakarta, NEM rata-rata EBTANAS mata pelajaran IPA Sekolah Lanjutan Tingkat Pertama (SLTP) di Kotamadia Surakarta tahun pelajaran 1996/1997 adalah masih rendah, yaitu 5,26 (Depdikbud, 1998).

Banyak kemungkinan yang menyebabkan rendahnya NEM mata pelajaran IPA tersebut. Kemungkinan pertama adalah karena para peserta mengalami kesulitan mengerjakan tes atau tingkat kesukaran perangkat tes tidak sesuai dengan kemampuan peserta tes. Kemungkinan kedua, karena tidak adanya sinkronisasi antara pendekatan pengajaran yang diterapkan dengan tipe soal. Sesuai dengan Garis-garis Besar Program Pengajaran (GBPP) tahun 1994 mata pelajaran IPA dilakukan dengan pendekatan keterampilan proses melalui metode Cara Belajar Siswa Aktif (CBSA). Kenyataannya soal-soal EBTANAS mata pelajaran IPA tidak mengarah kepada keterampilan proses, melainkan hanya kepada keterampilan hasil. Hal ini menunjukkan tingkat validitas yang belum memadai dari perangkat EBTANAS. Kemungkinan ketiga, NEM mata pelajaran IPA yang diperoleh siswa relatif homogen, sebagian besar memperoleh skor rendah. Hal ini menunjukkan bahwa NEM mata pelajaran IPA belum dapat memberikan informasi mengenai tingkat kemampuan siswa

yang sebenarnya. Dengan kata lain perangkat EBTANAS mata pelajaran IPA belum memiliki daya beda butir soal yang memadai. Kemungkinan keempat, karena waktu yang disediakan untuk penyusunan perangkat EBTANAS relatif terbatas (Depdikbud, 1997). Pengecoh yang dibuat kurang berfungsi sehingga penyebaran *option* jawaban tidak merata, mutu butir soal dan indeks reliabilitas perangkat EBTANAS mata pelajaran IPA rendah.

Penilaian mutu suatu perangkat tes semata-mata tidak hanya memperhatikan harga indeks reliabilitasnya. Indeks reliabilitas sangat berkaitan dengan kesalahan pengukuran. Indeks reliabilitas yang tinggi cenderung memiliki kesalahan pengukuran yang kecil. Gronlund (1982) mengemukakan bahwa kesalahan pengukuran memiliki kelebihan dibandingkan dengan indeks reliabilitas. Indeks reliabilitas dalam arti menunjukkan kualitas perangkat tes dengan membandingkan perbedaan antara dua hasil pengukuran atau lebih, sedangkan kesalahan pengukuran dapat digunakan untuk menaksir perolehan skor murni peserta tes.

Faktor-faktor kesalahan pengukuran bersumber dari seluruh komponen yang terlibat dalam proses pengukuran. Sumber-sumber kesalahan pengukuran, meliputi: perangkat tes, peserta tes, prosedur, penguji, dan lingkungan. Sumber kesalahan pengukuran menurut Nunnaly (1978) adalah variasi skor peserta tes, struktur sampel yang dipilih, dan variasi tingkat kesukaran butir soal. Berkaitan dengan mutu butir soal, Crocker dan Algina (1986) mengatakan bahwa apabila butir-butir soal yang membentuk perangkat tes tidak memiliki mutu yang memadai akan terjadi kesalahan pengukuran.

Tidak ada patokan tentang seberapa tinggi harga estimasi kesalahan pengukuran yang memuaskan untuk suatu perangkat tes. Diedrich (Gronlund, 1982) telah membuat Tabel yang memuat hubungan antara banyak butir soal dengan besarnya estimasi kesalahan pengukuran yang dapat ditolerir.

Tabel 1.
Tabel Pauld Diedrich
Hubungan Jumlah Butir Soal dan Estimasi Kesalahan Pengukuran

Jumlah Butir Soal	Estimasi Kesalahan Pengukuran
< 24	2
24 - 47	3
48 - 89	4
90 - 109	5
110 - 129	6
130 - 150	7

Teori pengukuran berkembang dari teori tes klasik ke teori respons butir. Keduanya bertujuan sama, yaitu: ingin mencari cara yang paling memuaskan untuk memperoleh skor yang benar-benar mencerminkan kemampuan peserta tes.

Menurut teori tes klasik, hasil pengukuran dapat dipercaya bila dalam beberapa kali pemberian tes terhadap sejumlah peserta tes diperoleh hasil yang relatif sama atau dengan kata lain hasil pengukuran tersebut reliabel. Indeks reliabilitas hasil pengukuran, selanjutnya hanya dituliskan sebagai indeks reliabilitas, dapat dipergunakan untuk memperkirakan tingkat kepercayaan hasil pengukuran. Berdasarkan indeks reliabilitas diperoleh informasi mengenai kecermatan hasil pengukuran sebagai estimasi skor sebenarnya yang dimiliki oleh peserta tes. Terdapat hubungan yang erat antara besarnya varians skor tampak dan varians kesalahan dengan indeks reliabilitas. Besarnya indeks reliabilitas digunakan untuk mengestimasi besarnya kesalahan pengukuran.

Ada beberapa model yang dapat digunakan untuk mengestimasi kesalahan pengukuran dalam teori tes klasik. Masing-masing model tersebut memiliki asumsi dan formulasi yang berbeda-beda. Pemilihan model tersebut tergantung pada asumsi yang dapat diterapkan pada seperangkat alat ukur yang akan diestimasi kesalahannya. Pemilihan model akan menentukan formulasi yang digunakan.

Kesalahan pengukuran selain dapat diestimasi dengan menggunakan teori tes klasik, juga dapat menggunakan teori respons butir. Pengukuran mutu perangkat tes pada teori respons butir menggunakan fungsi informasi perangkat tes yang merupakan jumlah dari semua fungsi informasi butir. Selanjutnya dapat ditentukan estimasi kesalahan pengukurannya yang berbanding terbalik dengan akar pangkat dua fungsi informasi perangkat tes tersebut (Hambleton et al., 1991: 37).

Perumusan masalah penelitian ini adalah

1. Berapa besar estimasi kesalahan pengukuran perangkat EBANAS mata pelajaran IPA SLTP di Kotamadia Surakarta tahun pelajaran 1997/1998 dengan menggunakan teori tes klasik ?
2. Berapa besar estimasi kesalahan pengukuran perangkat EBANAS mata pelajaran IPA SLTP di Kotamadia Surakarta tahun pelajaran 1997/1998 dengan menggunakan teori respons butir ?
3. Bagaimana hubungan besarnya estimasi kesalahan pengukuran teori tes klasik yang diperoleh dengan interval kepercayaan skor murni peserta EBANAS mata pelajaran IPA tahun pelajaran 1997/1998 di Kotamadia Surakarta ?
4. Bagaimana hubungan besarnya estimasi kesalahan pengukuran teori respons butir yang diperoleh dengan interval tingkat kemampuan peserta EBANAS mata pelajaran IPA tahun pelajaran 1997/1998 di Kotamadia Surakarta ?

Metode Penelitian

Populasi dan Sampel Penelitian

Populasi penelitian adalah seluruh lembar jawaban peserta EBANAS SLTP mata pelajaran IPA tahun pelajaran 1997/1998 di Kotamadia Surakarta untuk tipe soal obyektif pilihan ganda biasa. EBANAS tahun pelajaran 1997/1998 diikuti 92 SLTP / MTs.

Pelaksanaan EBANAS SLTP di Kotamadia Surakarta terdapat satu Rayon Penyelenggara yang dibagi menjadi tiga Subrayon, yaitu: Subrayon Barat, Subrayon Tengah, dan Subrayon Timur.

Paket Utama perangkat EBANAS SLTP mata pelajaran IPA yang digunakan hanya satu paket yang bernomor kode: 98 - C4 - U1. Selain itu juga

Estimasi Kesalahan Pengukuran EBTANAS Mata Pelajaran IPA SLTP di Kotamadya Surakarta

disediakan paket susulan perangkat EBTANAS SLTP mata pelajaran IPA yang bernomor kode: 98-C4 - S.

Jumlah seluruh lembar jawaban EBTANAS SLTP mata pelajaran IPA tahun pelajaran 1997/1998 di Kotamadia Surakarta sebanyak 13.681 lembar jawaban. Jumlah lembar jawaban Paket Utama perangkat EBTANAS 13.656 dan sisanya 25 lembar jawaban paket susulan perangkat EBTANAS.

Agar diperoleh sampel yang benar-benar mewakili populasi yang jumlahnya sangat besar, maka dilakukan pengambilan sampel dengan teknik *cluster stratified proportional systematic sampling*. Cluster dalam penelitian ini adalah SLTP/MTs yang ada di Kotamadia Surakarta yang siswa-siswanya mengikuti EBTANAS mata pelajaran IPA tahun pelajaran 1997/1998.

Berdasarkan perolehan rata-rata NEM IPA tahun pelajaran 1996/1997, SLTP/MTs yang ada di Kotamadia Surakarta dapat dikategorikan menjadi tiga tingkatan, yaitu: kelompok tinggi, kelompok menengah, dan kelompok rendah. Masing-masing kelompok SLTP/MTs diambil sampel secara sistematis dengan proporsi pengambilan 25 %. Sampel penelitian diperoleh 3.681 lembar jawaban.

Teknik Analisis Data

Perangkat soal yang baik harus teruji secara teoretis dan empiris. Perangkat tes yang baik secara teoretis harus memiliki validitas tinggi yang artinya perangkat tes tersebut telah mencakup materi yang seharusnya diukur. Mata pelajaran IPA SLTP mengacu pada GBPP tahun 1994 mata pelajaran IPA, sehingga validitas yang harus terpenuhi adalah validitas kurikulum. Kesesuaian perangkat EBTANAS mata pelajaran IPA dengan kurikulum dianalisis secara kualitatif.

Tinjauan mutu perangkat soal secara empiris ditentukan oleh karakteristik butir-butir soal yang meliputi: tingkat kesukaran butir soal, daya pembeda butir soal, dan efektivitas pengecoh melalui analisis kuantitatif menggunakan Program Iteman *version 3.00* dari *MicroCAT (tm) Testing System*. Tujuan analisis kuantitatif ini adalah untuk mengetahui mutu butir-butir soal perangkat EBTANAS mata pelajaran IPA. Berdasarkan kriteria yang telah dapat diketahui butir-butir soal yang diterima dan butir-butir soal yang ditolak.

Mengingat pada soal EBTANAS, bentuk pengukuran yang dapat dipenuhi adalah konjenerik dan Supri telah melakukan penelitian bahwa estimasi kesalahan pengukuran yang paling cermat adalah dengan menggunakan Model Feldt, maka pada penelitian ini untuk teori tes klasik digunakan Model Feldt.

Langkah-langkah yang dilakukan untuk mengestimasi kesalahan pengukuran dengan teori tes klasik model Feldt:

- a) Membelah perangkat EBTANAS menjadi dua bagian yang tidak harus sama panjang, asalkan isinya homogen. Homogenitas belahan didasarkan atas homogenitas tingkat kesukaran butir soal tiap belahan perangkat EBTANAS.
- b) Menyusun distribusi skor total tiap peserta tes, yaitu: jumlah respons jawaban benar tiap peserta EBTANAS.
- c) Menghitung indeks reliabilitas Feldt perangkat EBTANAS, digunakan persamaan:

$$r_{xx'} = \frac{4(S_{x_1x_2})}{S_x^2 \left[\frac{S_{x_1}^2 - S_{x_2}^2}{S_x} \right]^2}$$

Keterangan :

- $r_{xx'}$ = indeks reliabilitas perangkat tes
- $S_{x_1}^2$ = varians skor pada belahan pertama
- $S_{x_2}^2$ = varians skor pada belahan kedua
- $S_{x_1x_2}$ = kovarians skor pada belahan pertama dan kedua
- S_x = simpangan baku skor tes
- S_x^2 = varians skor tampak

- d) Menghitung varians skor murni menggunakan persamaan:

$$S_t^2 = S_x^2 \cdot r_{xx'}$$

- e) Menghitung varians kesalahan pengukuran perangkat EBTANAS menggunakan persamaan:

$$S_e^2 = S_x^2 - S_t^2$$

f) Menghitung besarnya kesalahan pengukuran perangkat EBTANAS:

$$KP = \sqrt{S_e^2}$$

Analisis empiris dengan teori respons butir menggunakan Program Ascal version 3.20 dari *MicroCAT (tm) Testing System*. Hasil analisis empiris diperoleh daya pembeda butir (a), tingkat kesukaran butir (b), dan peluang tebakan butir (c). Selain itu juga dapat diketahui harga kai kuadrat tiap butir soal dan tingkat kemampuan peserta EBTANAS. Selanjutnya dilakukan pengujian kecocokan data dengan model logistik tiga parameter.

Langkah-langkah estimasi kesalahan pengukuran dengan teori respons butir sebagai berikut:

- a) Melakukan pengujian kecocokan data dengan model logistik tiga parameter.
- b) Menghitung besarnya fungsi informasi perangkat EBTANAS dengan cara memasukkan harga a, b, dan c pada $\theta = -4,0; -3,5; -3,0; -2,5; -2,0; -1,5; -1,0; -0,5; 0; 0,5; 1,0; 1,5; 2,0; 2,5; 3,0; 3,5; \text{ dan } 4,0$ tiap-tiap butir soal yang cocok dengan model logistik tiga parameter ke dalam persamaan berikut :

$$I(\theta) = \sum_{j=1}^N \frac{(1-c_j)D^2 a_j^2}{\left[c_j + e^{D a_j (\theta - b_j)} \right] \left[1 + e^{-D a_j (\theta - b_j)} \right]^2}$$

(Naga, 1992)

Keterangan:

- θ = tingkat kemampuan peserta tes
- D = konstan yang berharga 1,7
- e = angka transendental yang berharga 2,718
- a = daya beda butir
- b = tingkat kesukaran butir
- c = peluang tebakan butir

- c) Menghitung besarnya kesalahan pengukuran (KP) perangkat EBANAS menggunakan persamaan :

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

Hasil Penelitian

Analisis data dengan menggunakan Program Iteman diperoleh butir-butir soal yang diterima dan ditolak berdasarkan klasifikasi daya pembeda, tingkat kesukaran, dan pengecoh butir soal. Berdasarkan klasifikasi butir-butir soal terdapat 28 butir soal yang diterima dan diikutkan dalam estimasi kesalahan pengukuran dengan teori tes klasik Model Feldt, yaitu: butir nomor 2, 3, 4, 7, 9, 10, 11, 12, 15, 16, 17, 21, 22, 24, 25, 26, 27, 28, 29, 30, 32, 33, 34, 37, 38, 40, 43, dan 45.

Perbandingan harga statistik hasil analisis perangkat EBANAS memperhatikan klasifikasi (perangkat EBANAS MK) dan perangkat EBANAS tanpa memperhatikan klasifikasi (TMK) sebagai berikut:

Tabel 2.
Rangkuman Indeks Reliabilitas Feldt, Varians Sekor Murni,
Varians Kesalahan Pengukuran, dan Estimasi Kesalahan Pengukuran

Hasil Analisis MK	Hasil Analisis TMK
$r_{xx'} = 0,7972$	$r_{xx'} = 0,8135$
$S_t^2 = 20,4138$	$S_t^2 = 37,2863$
$S_e^2 = 5,1941$	$S_e^2 = 8,5456$
KP = 2,2791	KP = 2,9233

Estimasi kesalahan pengukuran yang diperoleh dapat digunakan untuk mengestimasi interval sekor murni. Bila asumsi-asumsi yang mendasari interval kepercayaan sekor murni terpenuhi, maka interval kepercayaan sekor murni pada tingkat signifikansi $\alpha=0,05$ sbb.: $X - 4,4670 \leq T \leq X + 4,4670$

untuk interval kepercayaan sekor murni perangkat EBTANAS MK dan $X - 5,7297 \leq T \leq X + 5,7297$ untuk interval kepercayaan sekor murni perangkat EBTANAS TMK. X adalah sekor tampak peserta tes yang merupakan jumlah jawaban benar tiap peserta EBTANAS.

Butir-butir soal yang tidak cocok dengan teori respons butir model logistik tiga parameter ditetapkan berdasarkan *judgment* peneliti, yaitu: butir soal yang harga kaid kuadratnya berada di atas harga batas atas estimasi rerata kaid kuadrat. Penentuan kecocokan butir soal perangkat EBTANAS dengan model logistik tiga parameter menggunakan diperoleh butir soal yang cocok dengan model berjumlah 39 butir soal. Butir soal yang tidak cocok dengan model adalah butir soal nomor 16, 17, 19, 23, 38, dan 40.

Butir-butir soal perangkat EBTANAS yang dianalisis adalah butir-butir soal yang cocok dengan model logistik tiga parameter (perangkat BCM). Sebagai pembandingan juga dilakukan analisis semua butir-butir soal perangkat EBTANAS (perangkat SBE). Tabel 3 menampilkan besarnya estimasi kesalahan pengukuran EBTANAS tiap θ .

Tabel 3.
Estimasi Kesalahan Pengukuran Perangkat BCM dengan
Teori Respons Butir

No.	Harga θ	KP	No.	Harga θ	KP
1	-4,0	4,1883	10	0,5	0,3318
2	-3,5	2,8516	11	1,0	0,3348
3	-3,0	1,9596	12	1,5	0,3470
4	-2,5	1,3498	13	2,0	0,3669
5	-2,0	0,9358	14	2,5	0,4081
6	-1,5	0,6715	15	3,0	0,4784
7	-1,0	0,5344	16	3,5	0,5809
8	-0,5	0,4728	17	4,0	0,7260
9	0	0,3902			

Tabel 4.
Estimasi Kesalahan Pengukuran Perangkat SBE
dengan Teori Respons Butir

No.	Harga θ	KP	No.	Harga θ	KP
1	-4,0	3,3652	10	0,5	0,3192
2	-3,5	2,2128	11	1,0	0,3266
3	-3,0	1,4825	12	1,5	0,3413
4	-2,5	1,0197	13	2,0	0,3619
5	-2,0	0,7268	14	2,5	0,3791
6	-1,5	0,5460	15	3,0	0,3283
7	-1,0	0,4520	16	3,5	0,4303
8	-0,5	0,4137	17	4,0	0,6531
9	0,0	0,3614			

Estimasi kesalahan pengukuran terkecil tercapai pada $\theta = 0,5$. Interval tingkat kemampuan peserta EBANAS pada tingkat kepercayaan $\alpha = 0,05$; $\theta = 0,5$ dan $z_{\alpha/2} = 1,96$ adalah $-0,1504$ sampai dengan $1,1504$ untuk perangkat BCM dan $-0,1256$ sampai dengan $1,1256$ untuk perangkat SBE.

Pembahasan dan Kesimpulan

Estimasi kesalahan pengukuran dengan teori tes klasik Model Feldt

Hasil telaah cakupan materi perangkat EBANAS bentuk obyektif menunjukkan bahwa butir-butir soal yang dibuat telah memenuhi validitas isi kurikulum. Tapi bila memperhatikan pendekatan yang diterapkan, butir-butir soal tidak menunjukkan pendekatan keterampilan proses, melainkan masih pendekatan konsep. Pengamatan yang telah dilakukan peneliti di berbagai SLTP di Kotamadia Surakarta, sebenarnya masih banyak guru yang belum menerapkan pendekatan keterampilan proses secara murni. Kendala yang menyebabkannya adalah keterbatasan waktu dan atau tidak tersedianya sarana dan prasarana yang mendukung. Berdasarkan pengamatan tersebut dapat diambil kesimpulan bahwa butir-butir soal perangkat EBANAS yang diberikan tidak berdasarkan pendekatan keterampilan proses tidak memberikan

pengaruh yang besar terhadap kemungkinan terjadinya kesalahan pengukuran karena validitas kurikulum yang tidak sesuai.

Hasil estimasi kesalahan pengukuran perangkat EBTANAS TMK lebih tinggi daripada hasil estimasi kesalahan pengukuran perangkat EBTANAS MK. Meskipun harga indeks reliabilitas perangkat EBTANAS MK lebih rendah dibandingkan dengan harga indeks reliabilitas perangkat EBTANAS TMK. Harga indeks reliabilitas dalam arti stabilitas yang tinggi tidak menjamin perolehan estimasi kesalahan pengukuran menjadi rendah. Tinggi rendahnya estimasi kesalahan pengukuran sangat tergantung pada varians sekor murninya.

Perangkat EBTANAS MK telah bebas dari karakteristik butir soal yang menyebabkan rendahnya mutu perangkat EBTANAS, sehingga estimasi kesalahan pengukurannya lebih rendah. Semakin kecil perolehan harga estimasi kesalahan pengukuran, maka pengukuran semakin cermat dan dapat dipercaya.

Bila membandingkan perolehan harga estimasi kesalahan pengukuran baik estimasi kesalahan pengukuran perangkat EBTANAS MK maupun estimasi kesalahan pengukuran perangkat EBTANAS TMK dengan Tabel 1 menunjukkan kesesuaian antara jumlah butir soal dengan besarnya harga estimasi kesalahan pengukuran yang dapat ditolerir. Jumlah butir soal perangkat EBTANAS MK adalah 28, estimasi kesalahan pengukurannya adalah 2,2791. Jumlah butir soal perangkat EBTANAS TMK adalah 45, estimasi kesalahan pengukurannya adalah 2,9233. Berdasarkan Tabel tersebut perangkat soal yang memiliki jumlah butir soal antara 24 sampai 47, estimasi kesalahan pengukuran maksimum adalah 3.

Interval sekor murni perangkat EBTANAS MK lebih sempit daripada interval kepercayaan sekor murni perangkat EBTANAS TMK. Interval kepercayaan sekor murni perangkat EBTANAS MK dan interval kepercayaan sekor murni perangkat EBTANAS TMK luasnya adalah 8,9340 angka dan 11,4594 angka. Semakin kecil estimasi kesalahan pengukuran, semakin sempit interval kepercayaan sekor murni yang berarti hasil pengukuran semakin cermat.

Estimasi kesalahan pengukuran dengan teori respons butir

Estimasi kesalahan pengukuran minimum terjadi pada saat fungsi informasi maksimum, yaitu pada $\theta = 0,5$. Ini menunjukkan bahwa perangkat EBTANAS tersebut cocok untuk peserta tes yang memiliki tingkat kemampuan θ sekitar 0,5. Penafsiran hasil estimasi kesalahan pengukuran tidak dapat secara langsung, tapi harus dibuat dalam suatu interval tingkat kemampuan peserta EBTANAS.

Interval tingkat kemampuan peserta EBTANAS perangkat BCM lebih lebar dibandingkan dengan interval tingkat kemampuan peserta EBTANAS perangkat SBE. Hal tersebut wajar terjadi karena estimasi kesalahan pengukuran berbanding terbalik dengan fungsi informasi perangkat tes, sehingga semakin besar harga fungsi informasi maka semakin rendah estimasi kesalahan pengukurannya. Harga fungsi informasi perangkat tes merupakan penjumlahan dari harga fungsi informasi butir-butirnya. Perangkat BCM telah mengalami pengurangan jumlah butir, sehingga fungsi informasinya berkurang pula.

Semakin lebar interval tingkat kemampuan, ketidakpastian tentang tingkat kemampuan peserta tes yang sebenarnya semakin tinggi. Estimasi kesalahan pengukuran perangkat SBE lebih rendah daripada estimasi kesalahan pengukuran perangkat BCM. Analisis data perangkat SBE masih mengikutsertakan butir-butir soal yang tidak cocok dengan model logistik tiga parameter, sehingga hasil analisis parameter butirnya tidak akurat.

Harga fungsi informasi perangkat BCM dan perangkat SBE termasuk kriteria sangat tinggi (Hambleton et al., 1991). Perangkat BCM pada tingkat kemampuan peserta $\theta = -1,0$ sampai 3,0 dan perangkat SBE pada tingkat kemampuan $\theta = -1,5$ sampai dengan 3,5. Hal ini menunjukkan bahwa butir-butir soal perangkat EBTANAS mampu mengungkap informasi mengenai parameter tingkat kemampuan peserta EBTANAS untuk kelompok menengah hingga kelompok tinggi. Butir-butir soal perangkat EBTANAS telah memiliki daya beda yang cukup memadai, karena telah dapat membedakan kelompok rendah dengan kelompok menengah hingga kelompok tinggi.

Berdasarkan uraian di atas dapat disimpulkan mengenai perangkat EBTANAS mata pelajaran IPA SLTP di Kotamadia Surakarta tahun pelajaran 1997/1998 sebagai berikut:

1. Secara empiris telah terbukti bahwa menurut teori tes klasik, estimasi kesalahan pengukuran perangkat EBTANAS memperhatikan klasifikasi lebih rendah dibandingkan dengan estimasi kesalahan pengukuran perangkat EBTANAS tanpa memperhatikan klasifikasi. Harga estimasi kesalahan pengukuran perangkat EBTANAS memperhatikan klasifikasi adalah 2,2791 dan harga estimasi kesalahan pengukuran perangkat EBTANAS tanpa memperhatikan klasifikasi adalah 2,9233.
2. Perangkat EBTANAS dengan memperhatikan klasifikasi dan perangkat EBTANAS tanpa memperhatikan klasifikasi memiliki estimasi kesalahan pengukuran yang masih dapat ditolerir. Estimasi kesalahan pengukuran untuk 24 sampai 47 butir soal maksimal 3.
3. Interval kepercayaan sekor murni peserta EBTANAS perangkat EBTANAS memperhatikan klasifikasi pada tingkat kepercayaan $\alpha = 0,05$ adalah $X - 4,4670 \leq T \leq X + 4,4670$ dan interval kepercayaan sekor murni peserta EBTANAS perangkat EBTANAS tanpa memperhatikan klasifikasi pada tingkat kepercayaan $\alpha = 0,05$ adalah $X - 5,7297 \leq T \leq X + 5,7297$.
4. Secara empiris estimasi kesalahan pengukuran menurut teori respons butir, baik hasil analisis semua butir soal perangkat EBTANAS maupun hasil analisis butir soal yang cocok dengan model logistik tiga parameter menunjukkan estimasi kesalahan pengukuran paling rendah terjadi pada tingkat kemampuan $\theta = 0,5$.
5. Interval kepercayaan tingkat kemampuan peserta EBTANAS perangkat EBTANAS yang butir-butir soalnya cocok dengan model logistik tiga parameter pada tingkat kepercayaan $\alpha = 0,05$ adalah $-0,1504$ sampai dengan $1,1504$ dan interval kepercayaan tingkat kemampuan peserta EBTANAS perangkat EBTANAS yang semua butir soalnya dianalisis pada tingkat kepercayaan $\alpha = 0,05$ adalah $-0,1256$ sampai dengan $1,1256$.

Daftar Pustaka

- Allen, M.J dan Yen, W.M. (1979). *Introduction to Measurement Theory*. California: Brooks/Cole Publishing Company.
- Azwar, Saifuddin. (1997). *Reliabilitas dan Validitas edisi ke-3*. Yogyakarta: Pustaka Pelajar.
- Crocker, L dan Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehard and Winston Inc.
- Depdikbud. (1994). *Garis-garis Besar Program Pengajaran (GBPP) IPA SLTP*. Jakarta: Dikdasmen.
- Depdikbud. (1997). *Laporan Pelaksanaan Evaluasi Belajar Tahap Akhir Nasional (EBTANAS) Tahun Pelajaran 1996/1997*. Surakarta: Kandepdikbud.
- Depdikbud, Dinas P&K, Depag. (1998). *Keputusan Bersama Kepala Kantor wilayah Departemen Pendidikan dan Kebudayaan Propinsi Jawa Tengah dan Kepala Dinas Pendidikan dan Kebudayaan Propinsi Dati I Jawa Tengah serta Kepala Kantor Wilayah Departemen Agama Propinsi Jawa Tengah Nomor: 0083/I03.02/MN/1998; 423.7/01296/1998; Wk/5.a/PP.01.1/696/1998 tanggal 6 Maret 1998*. Semarang: Depdikbud-Dinas P&K-Depag.
- Depdikbud. (1998). *Rekapitulasi Peserta dan Pengolahan Evaluasi Belajar Tahap Akhir Nasional (EBTANAS) Tahun Pelajaran 1997/1998*. Semarang: Kanwil Depdikbud Propinsi Jawa Tengah bekerjasama dengan PT Wahana komputer.
- Feldt, L.S. (1984). "Some Relationships Between The Binomial Error Model and Classical Test Theory". *Educational and Psychological Measurement*. Vol. 44, No. 4, Pp. 883-891.

Estimasi Kesalahan Pengukuran EBTANAS Mata Pelajaran IPA SLTP di Kotamadya Surakarta

- Gronlund, N.E. (1982). *Constructing achievement test*. (Rev. ed). Englewood Cliffs, N.J: Prentice Hall.
- Hambleton, R.K et al. (1991). *Fundamental of Item Response Theory*. Newbury Park London New Delhi: Sage Publications.
- Linn, R.L. (1989). *Educational Measurement* 3rd ed. London: Collier Mcmillan Publisher.
- Naga, Dali S. (1992). *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Nunnaly, J.C. (1978). *Psychometric Theory* 2nd ed. New York: McGraw - Hill Book Company.
- Sudjana. (1992). *Metoda Statistika* edisi ke-5. Bandung: Penerbit Tarsito.
- Supi'i, Imam. (1997). *Estimasi Kesalahan Pengukuran Tes Bahasa Inggris dengan Model Binomial, Thorndike, Kristof, dan Feldt*. Tesis. Universitas Gajah Mada.