

KARAKTERISTIK SOAL TES EBTRANAS SEKOLAH DASAR DI KABUPATEN JEPARA

*Oleh:
Saptariana dan Sumarno*

Abstrak

Penelitian ini bertujuan untuk memperoleh gambaran tentang karakteristik soal dan kemampuan peserta tes Ebtanas SD di Kabupaten Jepara. Populasi penelitian ini adalah soal dan seluruh lembar jawaban siswa pada mata pelajaran Bahasa Indonesia, Matematika dan Ilmu Pengetahuan Sosial (IPS). Data penelitian dianalisis secara kualitatif dan kuantitatif dengan menggunakan estimasi parameter butir program Rascal. Hasil analisis data menunjukkan: ketiga perangkat tes tersebut untuk bagian tes objektif maupun bagian tes esainya menunjukkan sebagai perangkat tes yang berkualitas baik. Nilai fungsi informasi butir pada masing-masing perangkat tes nilainya tidak jauh berbeda, yang berarti kualitas ketiga perangkat tes Ebtanas tersebut memiliki tingkat reliabilitas yang hampir sama. Selain itu daya pembeda pada semua perangkat tes, semuanya menunjukkan sebagai butir-butir soal yang baik karena dapat membedakan antara mereka yang menguasai materi dengan yang tidak menguasai materi.

Pendahuluan

Tes yang digunakan untuk evaluasi hasil belajar pada tingkat akhir pada jenjang SD, SLTP maupun SLTA yang diselenggarakan secara nasional disebut dengan Evaluasi Belajar Tahap Akhir Nasional (Ebtanas) dengan hasil berupa Nilai Ebtanas Murni (NEM).

NEM digunakan untuk proses seleksi masuk pada jenjang pendidikan yang lebih tinggi. Selain itu hasil dari Ebtanas juga digunakan untuk memantau mutu pendidikan, yaitu untuk melihat prestasi yang dicapai oleh

Karakteristik Soal Tes Ebtanas Sekolah Dasar di Kabupaten Jepara

siswa dari materi-materi yang telah diberikan. Mengingat pentingnya hasil Ebtanas, maka tes Ebtanas itu harus memiliki standar tes yang baik, sehingga tidak ada pihak-pihak yang dirugikan.

Ada dua hal yang harus diperhatikan dalam pembuatan tes. *Pertama*, tes tersebut harus dapat mengungkap secara benar ciri terpendam yang ada pada peserta tes, yang berhubungan dengan validitas pada konstruksi uji tes. *Kedua*, tes tersebut harus dapat mengungkapkan tanggapan dari peserta tes dengan informasi yang dapat dipercaya, yang berhubungan dengan reliabilitas pada uji tes (Naga, 1992).

Untuk Menguji soal/tes tersebut baik atau tidak maka harus dilakukan pengujian secara teoritik (penelaahan soal) dan pengujian secara empirik (Suryabrata, 1987). Pengujian soal secara teoritik mempunyai tiga sasaran yaitu (a) kesesuaian isi soal dengan hal-hal yang akan diuji (b) kesesuaian soal-soal itu dengan syarat-syarat psikometris dan (c) ketepatan dan kecermatan rumusan soal-soal tersebut. Sedangkan pengujian secara empiris adalah melakukan analisis soal secara kuantitatif.

Model pengukuran pendidikan untuk menganalisis tes hasil belajar dapat menggunakan model pengukuran klasik dan model pengukuran modern (Naga, 1992: 4). Pengukuran klasik mempunyai ciri, kelompok butir pada tes tidak dapat dipisahkan dari kelompok peserta, maksudnya adalah jika kelompok butir tes yang sama ditempuh oleh kelompok peserta yang berbeda, maka ciri atau karakteristik kelompok butir itu pada umumnya berubah. Sedang pengukuran modern mempunyai ciri melepaskan ketakterpisahan di antara butir tes dan peserta tes.

Pada penelitian ini untuk mengetahui bagaimana karakteristik tes Ebtanas SD, atau kualitas dari tes Ebtanas, maka dilakukan pengukuran secara empirik dengan menggunakan model pengukuran modern (model pengukuran teori respons butir), khususnya model pengukuran satu parameter (*Rasch Model*).

Untuk memantau kualitas tes, analisis data dilakukan terhadap tes pilihan ganda (tes objektif) serta tes esai dengan bidang studi yang diambil dibatasi pada bidang studi matematika, Bahasa Indonesia dan IPS pada Ebtanas SD tahun ajaran 1997/1998.

Permasalahan yang ada pada penelitian Ebtanas SD tahun ajaran 1997/1998 di Kabupaten Jepara adalah: berapa besar parameter butir soal-soal

pilihan ganda, bagaimana tingkat kemampuan peserta dalam menjawab tes pilihan ganda, butir-butir manakah yang cocok dengan model *Rasch* pada tes pilihan ganda, butir-butir manakah yang memberikan fungsi informasi butir maksimum pada tes pilihan ganda, serta berapa besar parameter butir soal-soal esai.

Tujuan yang akan dicapai dalam penelitian Ebtanas SD tahun ajaran 1997/1998 di Kabupaten Jepara ini adalah: ingin memperoleh informasi tentang parameter butir soal-soal pilihan ganda, ingin mengetahui tingkat kemampuan peserta dalam menjawab tes pilihan ganda, ingin mengetahui butir-butir yang cocok dengan model *Rasch* pada tes pilihan ganda, ingin mengetahui fungsi informasi butir maksimum pada tes pilihan ganda, serta ingin mengetahui parameter butir soal-soal esai.

Pengukuran dan Penilaian

Penilaian sangat erat kaitannya dengan pengukuran karena sebelum adanya penilaian harus ada pengukuran lebih dahulu. Ratna Sayekti (1988: 5) berpendapat bahwa pengukuran adalah prosedur pemberian angka atau nilai pada diri orang-orang sedemikian rupa sehingga hubungan antara orang-orang itu dengan ciri di dalamnya tetap ada sesuai keadaan yang sebenarnya. Hal ini sejalan dengan definisi dari Cronbach yang dikutip oleh Mehrens dan Lehman (1973: 6) bahwa pengukuran merupakan suatu prosedur yang sistematis untuk mengamati perilaku seseorang dan menggambarkannya dengan bantuan skala numerik atau sistem pengkategorian.

Setelah diadakan pengukuran terhadap tes, proses selanjutnya diadakan penilaian yang tujuannya untuk mengambil suatu keputusan baik atau buruk. Penilaian, menurut Mehrens dan Lehman (1973: 6) adalah merupakan proses penggambaran, memperoleh dan memberikan informasi yang berguna sebagai alternatif pengambilan keputusan. Sedangkan menurut Umar (1989) penilaian adalah kegiatan untuk mengetahui apakah suatu program telah berhasil dan efisien.

Evaluasi atau penilaian di bidang pendidikan khususnya evaluasi terhadap prestasi belajar siswa sebagian besar bersumber dari hasil-hasil pengukuran. Oleh karena itu masalah pengukuran mempunyai kedudukan yang sangat penting di dalam proses evaluasi. Baik buruknya evaluasi akan banyak bergantung pada hasil-hasil pengukuran yang dilakukan, sehingga

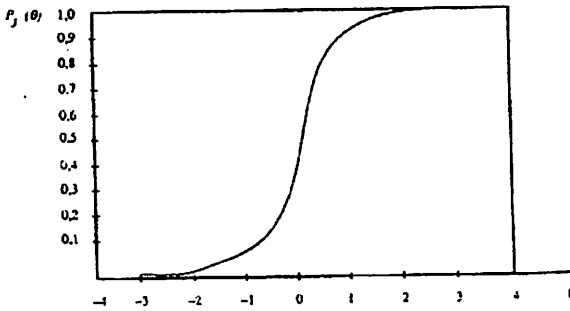
hasil pengukuran yang kurang cermat akan memberikan hasil evaluasi yang kurang cermat pula.

Konsep Dasar Teori Respons butir (*Item Response Theory: IRT*)

Hal yang mendasar dari IRT menurut Hambleton, dkk (1991: 7) adalah *pertama*, performansi seorang subjek pada suatu butir dapat diprediksi oleh seperangkat faktor yang disebut *trait* (kemampuan) yang dimilikinya. *Kedua*, hubungan antara kemampuan subyek pada suatu butir (hasil) tes dan kemampuan (*ability*) yang mendasarinya dapat digambarkan oleh suatu fungsi yang menaik secara monoton yang disebut kurve karakteristik item (*item characteristic curve* ICC). Sedangkan menurut Hulin (1983: 14) bagian yang mendasar dari IRT adalah fungsi matematika dan hubungan probabilitas respon terhadap item oleh individu untuk menentukan karakteristik dari individu dan item. Berdasar kedua pendapat tersebut maka dapat dikatakan bahwa di dalam teori respons butir yang dibicarakan adalah hubungan respon individu terhadap item dengan kemampuan individu.

Menurut Naga (1992: 6) parameter butir terdiri dari tingkat kesukaran butir, daya pembeda dan faktor kebetulan. Sedangkan menurut Lord (1980: 12) parameter butir terdiri dari: a, b, c, dimana a sebagai daya beda item, b sebagai tingkat kesukaran dan c sebagai *guessing*.

Hubungan antara respon terhadap item dengan kemampuan individu digambarkan dalam bentuk kurve yang disebut *item characteristic curve* (ICC). Dari kurve ini dapat dilihat bahwa peserta tes dengan kemampuan tinggi akan memiliki probabilitas yang lebih besar untuk menjawab benar terhadap suatu item sedangkan peserta dengan kemampuan rendah akan memiliki probabilitas yang kecil untuk menjawab benar terhadap suatu item. Jadi dari kurve ini dapat diketahui probabilitas seseorang yang memiliki kemampuan tertentu dalam menjawab suatu butir soal. Gambar dari model ICC tersebut adalah (Naga, 1992: 163):



Gambar 1. Kurva Karakteristik Butir

Model Teori Respons Butir

Teori respons butir mempunyai formula matematika yang menghubungkan ciri peserta dan ciri butir yang disebut model. Ciri butir dan ciri peserta yang dihubungkan oleh model membentuk fungsi atau lengkungan grafik yang disebut ICC (*Item Characteristic Curve*). Menurut Naga (1992) model-model dalam teori respons butir adalah model skala sempurna, model jarak laten, model ogive normal dan model logistik. Model logistik banyak digunakan karena analisis persamaan matematikanya sederhana. Model logistik ini memiliki tiga model yaitu model logistik tiga parameter, dua parameter dan satu parameter, yang pada penelitian ini analisis dilakukan berdasar model logistik satu parameter. Karena analisis hanya didasarkan model logistik satu parameter maka hanya akan dibahas teori model logistik satu parameter.

Sesuai dengan namanya model logistik satu parameter ini memiliki satu parameter butir yaitu parameter tingkat kesukaran (b). Sedangkan parameter daya beda (a) dan parameter dugaan atau *pseudo* (c) sama dengan nol. Model ini diajukan oleh George Rasch (1960) yang dinamakan dengan karakteristik khusus dari fungsi logistik sederhana dimana pengukurannya dilakukan secara umum dan memungkinkan adanya obyektifitas spesifik. Persamaan model Rasch adalah (Wright dan Stone, 1979: 15)

$$P \{X_{1i} = 1 | \beta_i, -\delta_i\} = \exp(\beta_i - \delta_i) / [1 + \exp(\beta_i - \delta_i)] \dots\dots (1)$$

dimana $P \{ X_{it} \}$ adalah peluang menjawab benar subyek v pada item t . $X_{it} = 1$ bila jawaban benar dan $X_{it} = 0$ bila jawaban salah. β_j : tingkat kemampuan dan δ_t merupakan tingkat kesukaran butir dan konstanta eksponensial dengan nilai 2,718.

Model *Rasch* menganjurkan di dalam mengestimasi β_j dan δ_t harus dilepaskan keterikatan antara satu dengan yang lain misalnya jika mengestimasi β_j maka harus dibebaskan dari pengaruh δ_t demikian juga halnya jika mengestimasi δ_t maka harus dibebaskan dari pengaruh β_j . Parameter β_j digunakan untuk mengukur kemampuan peserta sedangkan parameter δ_t untuk mengkalibrasi kesukaran butir soal. Di dalam mengestimasi β_j dan δ_t akan diperoleh nilai maksimum dari keadaan *likelihood* sehingga hasilnya tidak bias, tidak berubah, efisien dan baik. Satuan untuk tingkat kemampuan dan tingkat kesukaran dalam *Rasch* model dinamakan logits.

Asumsi yang digunakan pada model *Rasch* menurut Djemari Mardapi (1991: 12) ada dua yaitu: (1) semua butir memiliki daya pembeda yang sama dan (2) peluang menjawab butir benar bagi mereka yang memiliki kemampuan rendah sama dengan nol. Jadi dari asumsi ini dapat dikatakan semua kurve karakteristik butir-butir model ini adalah sejajar dan mendekati sejajar, karena besarnya slop kurve sama dan memotong sumbu y di titik 0 atau mendekati titik 0.

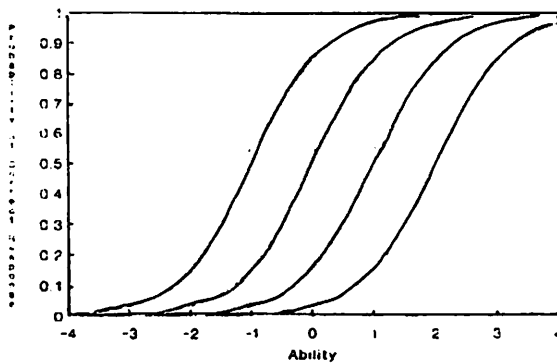
Model logistik satu parameter sering dinamakan model *Rasch* karena secara matematika ekuivalen dengan model *Rasch* (Hambleton, dkk, 1991: 14). yang mempunyai persamaan:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad i = 1, 2, 3 \dots n \dots \dots \dots (2)$$

Parameter b_i pada suatu butir adalah batasan skala kemampuan dimana peluang menjawab benar adalah 0,5. Parameter ini merupakan parameter lokasi, yang menunjukkan letak posisinya di dalam ICC pada skala kemampuan. Besarnya nilai parameter b_i menunjukkan besarnya kemampuan yang dibutuhkan untuk menjawab butir benar 50 %. Pada ICC, jika soal sulit skala tingkat kesukaran terletak di sebelah kanan atau bagian atas dari skala

kemampuan sedangkan bila soal mudah tingkat kesukaran terletak di sebelah kiri atau bagian bawah dari skala kemampuan.

Jika nilai dari kemampuan suatu group ditransformasikan maka mean = 0 dan standar deviasi = 1. Dan nilai b_i bervariasi dari -2 dan 2. Nilai b_i yang dekat dengan -2 menunjukkan butir yang mudah dan nilai b_i yang dekat 2 menunjukkan butir yang sukar. Kurve karakteristik butir untuk satu parameter adalah seperti gambar 2. di bawah ini (Hambleton, dkk, 1991: 14):



Gambar 2. Kurve Karakteristik Butir Satu Paramater

Metode dan Analisis Data Penelitian

Penelitian dilakukan di wilayah kabupaten Jepara dan dilaksanakan setelah berlangsungnya Ebtanas SD tahun ajaran 1997/1998. Populasi dalam penelitian ini adalah soal dan seluruh lembar jawaban siswa peserta tes Ebtanas Sekolah Dasar tahun ajaran 1997/1998 se-Kabupaten Jepara pada mata pelajaran bahasa Indonesia, matematika dan Ilmu Pengetahuan Sosial.

Untuk menentukan sampelnya digunakan teknik *two stage cluster sampling*. *Stage* pertama untuk menentukan pencuplikan sekolah dan *stage* kedua untuk menentukan pencuplikan siswa. Pada pencuplikan sekolah dipilih secara acak dari sekolah-sekolah dasar yang ada di Kabupaten Jepara, sedangkan sampel siswa diambil dari semua siswa yang ikut Ebtanas dari sekolah yang terpilih.

Karakteristik Soal Tes Ebtanas Sekolah Dasar di Kabupaten Jepara

Naga (1992: 299) mengatakan bahwa untuk model satu parameter (1P), pengestimasi parameter harus didasarkan pada lembar jawaban tes yang dikerjakan tidak boleh kurang dari 500 peserta. Berdasarkan pada aturan dan batasan minimal di atas, dalam penelitian ini sampel yang didapat adalah 633 lembar jawab siswa untuk masing-masing pelajaran: bahasa Indonesia, IPS dan matematika.

Analisis data untuk tes pilihan ganda dilakukan dengan dua pendekatan, yaitu pendekatan kualitatif dan pendekatan kuantitatif. Pendekatan kualitatif berupa analisis data melalui penelaahan butir-butir soal pada setiap perangkat tes. Penelaahan ini dilakukan oleh tiga orang penelaah ahli atau *expert* yang dilakukan secara terpisah. Cara penelaahan ditentukan berdasar pedoman telaah butir soal yang dikeluarkan oleh Puslitbang Sisjian Depdiknas.

Analisis data berikutnya adalah analisis dengan pendekatan kuantitatif. Data yang dianalisis berupa respons peserta Ebtanas SD tahun 1997/1998 di Kabupaten Jepara pada perangkat tes bahasa Indonesia, Matematika dan IPS. Pada pendekatan ini analisis data dilakukan beberapa tahap yang meliputi tahapan: menentukan karakteristik butir soal, kecocokan data setiap butir soal terhadap model dan estimasi tingkat kemampuan peserta. Pada pendekatan secara kuantitatif, respons peserta Ebtanas dianalisis menggunakan program komputer RASCAL™ Versi 3.00 tahun 1988 dari MicroCAT™ Testing System pada setiap perangkat tes.

Berdasarkan hasil analisis, dapat diketahui butir soal yang mempunyai karakteristik tingkat kesukaran tinggi, normal dan rendah setelah melihat harga parameter tingkat kesukaran butir (b_i). Butir soal bisa dikategorikan sebagai butir soal dengan tingkat kesukaran normal apabila $-2 \leq b_i \leq 2$, dikategorikan sebagai butir soal yang sukar apabila $b_i > 2$ dan butir soal dikategorikan sebagai butir soal yang mudah apabila $b_i < -2$ (Hambleton, dkk., 1991: 13).

Untuk mengetahui kecocokan butir terhadap model dilihat berdasarkan nilai *Chi-Square* (χ^2) yang diperoleh dan dibandingkan dengan nilai *Chi-Square* (χ^2) tabel. Jika nilai *Chi-Square* (χ^2) hitung lebih besar dari nilai *Chi-Square* (χ^2) tabel maka berarti butir soal tidak cocok dengan model. Jika nilai *Chi-Square* (χ^2) hitung lebih kecil dari *Chi-Square* (χ^2) tabel berarti butir tersebut cocok dengan model.

Untuk mengetahui tingkat kemampuan peserta dapat dilihat dari skor peserta pada hasil analisis *print-out* program *Rascal*. Dalam penelitian ini tingkat kemampuan peserta dikategorikan dalam 3 kelompok yaitu: tingkat kemampuan rendah (skor peserta < -1), tingkat kemampuan sedang ($-1 \leq$ skor peserta ≤ 1) dan tingkat kemampuan tinggi (skor peserta > 1).

Untuk mengetahui fungsi informasi butir dilakukan dengan langkah-langkah sebagai berikut: (1) tentukan tingkat kemampuan peserta dari -4 s.d. 5 dengan jarak 0,5. (2) masukkan nilai-nilai a dan b hasil *print-out* program *Rascal* ke dalam rumus. (3) untuk menentukan fungsi informasi butir maksimum didasarkan atas teori yang dikemukakan Birnbaum, bahwa fungsi informasi maksimum berasal dari θ maksimum ($\theta_{maks.}$), dimana jika nilai $c = 0$ maka $\theta_{maks.} = b$.

Hasil Analisis

Hasil analisis kualitatif soal yang didapat dari penelaah ahli dapat dipilah menjadi dua kategori butir soal, yaitu butir soal yang diterima tanpa revisi dan butir soal yang diterima dengan revisi. Suatu butir soal dikategorikan diterima tanpa revisi apabila semua kriteria yang terdapat pada setiap bidang telaah kualitatif yaitu bidang: materi, konstruksi dan bahasa telah terpenuhi, sedangkan suatu butir soal dikategorikan diterima dengan revisi apabila minimal ada satu kriteria dari ketiga bidang telaah tersebut (materi, konstruksi dan bahasa) tidak terpenuhi. Dan apabila seluruh kriteria tidak terpenuhi pada suatu butir soal, berarti butir soal tersebut termasuk butir soal yang jelek atau tidak diterima.

Dari hasil analisis, diketahui soal-soal pilihan ganda pada perangkat tes Bahasa Indonesia yang perlu direvisi sebanyak 2 butir soal. Untuk soal-soal perangkat tes Matematika yang memerlukan perbaikan sebanyak 11 butir soal. Sedangkan soal-soal pada perangkat tes IPS yang tergolong memerlukan perbaikan sebanyak 3 butir soal.

Analisis secara kuantitatif (empiris) dilakukan untuk mengetahui tingkat kesukaran butir, kecocokan butir terhadap model dan karakteristik peserta tes. Untuk menganalisis hal tersebut di atas dilakukan menggunakan komputer dengan paket program *Rascal*.

Karakteristik Soal Tes Ebtanas Sekolah Dasar di Kabupaten Jepara

Hasil analisis menyatakan butir soal diterima maksudnya butir tersebut dapat menggambarkan kemampuan sebenarnya yang dimiliki siswa, karena siswa yang berkemampuan tinggi dapat menjawab soal dengan benar, sedangkan siswa yang berkemampuan rendah tidak bisa menjawab dengan benar.

Berdasar hasil analisis Rascal tentang parameter tingkat kesukaran akan terlihat butir soal yang baik dan butir soal yang tidak baik. Hasil analisis menunjukkan bahwa pada perangkat tes bahasa Indonesia ada 25 soal yang diterima. Pada perangkat tes Matematika ada 28 soal yang diterima, sedangkan soal yang dikategorikan sulit untuk perangkat tes Matematika tidak ada. Pada perangkat tes Ilmu Pengetahuan Sosial (IPS) ada 28 soal yang diterima.

Analisis yang berkaitan dengan kecocokan butir terhadap model didasarkan atas model *Rasch*. Kecocokan butir soal terhadap model *Rasch* dilihat berdasarkan harga *Chi square* (χ^2) yang diperoleh, dan dibandingkan dengan harga *Chi square* (χ^2) dalam tabel berdasarkan hasil *print-out* program Rascal.

Berdasar perbandingan antara nilai χ^2 diketahui bahwa butir soal yang **cocok** dengan model *Rasch* pada pelajaran Bahasa Indonesia ada sejumlah 25 soal, sedangkan butir yang **tidak cocok** dengan model *Rasch* sejumlah 10 soal. Pada mata pelajaran Matematika, butir soal yang **cocok** dengan model *Rasch* ada sejumlah 20 soal, sedangkan butir soal yang **tidak cocok** dengan model *Rasch* sejumlah 10 soal. Untuk mata pelajaran IPS, butir soal yang **cocok** dengan model *Rasch* ada sejumlah 25 soal, sedangkan butir yang **tidak cocok** dengan model *Rasch* sejumlah 10 soal.

Tingkat kemampuan peserta dalam teori respons butir dinyatakan dalam skala logits. Secara teoretis, nilai baku tingkat kemampuan dalam teori respons butir membentang dari minus tak terhingga sampai plus tak terhingga. Tetapi secara praktis, nilai baku yang berguna hanya terletak antara -4 sampai +4 atau ada kalanya yang menetapkan nilai baku terletak diantara -3 sampai +3. Di dalam analisis *Rasch* menggunakan paket program Rascal, diasumsikan bahwa kemampuan (*ability*) peserta tes berdistribusi normal dan memiliki rerata 0 serta varians 1, atau $N(0,1)$.

Pada perangkat tes Bahasa Indonesia tingkat kemampuan peserta berada pada nilai -2,454 hingga 2,913. Pada perangkat tes Matematika tingkat kemampuan peserta berada pada nilai -2.505 hingga 2,796. Pada perangkat tes IPS tingkat kemampuan peserta dari -4,302 hingga 3,369. Dalam penelitian ini tingkat kemampuan peserta dikategorikan dalam 3 kelompok yaitu tingkat kemampuan rendah (lebih kecil dari -1), tingkat kemampuan sedang (antara -1 sampai dengan 1) dan tingkat kemampuan tinggi (lebih besar dari 1).

Hasil informasi butir menunjukkan bahwa untuk ketiga perangkat tes Ebtanas memberikan nilai fungsi informasi butir dari 0,1 sampai dengan 1,4. Nilai fungsi informasi tersebut tersebar dan bervariasi pada skala tingkat kemampuan mulai dari skala terendah -4 sampai dengan +4 dengan interval 0,5, yang selanjutnya diambil nilai fungsi informasi yang maksimum. Penentuan nilai fungsi informasi maksimum adalah sesuai dengan pendapat Birnbaun, bahwa fungsi informasi maksimum berasal dari θ_{maksimum} , dimana jika $c = 0$ maka $\theta_{\text{maksimum}} = b$.

Pada perangkat tes Bahasa Indonesia diketahui nilai fungsi informasi maksimum berada pada nilai 0,198... sampai dengan 0,200.... Nilai fungsi informasi butir yang paling besar yaitu 0,200657 pada butir soal nomor: 1, 2 dan 3, sedangkan fungsi informasi butir yang paling kecil sebesar 0,198151 pada butir soal nomor 8.

Untuk perangkat tes Matematika nilai fungsi informasi maksimum berada pada nilai 0,057... sampai dengan 0,336.... Nilai fungsi informasi butir yang paling besar yaitu 0,336051 pada butir soal nomor 23, sedangkan fungsi informasi butir yang paling kecil sebesar 0,057655 pada butir soal nomor 29. Sedangkan pada perangkat tes Ilmu Pengetahuan Sosial diketahui nilai fungsi informasi maksimum berada pada nilai 0,106... sampai dengan 1,474.... Nilai fungsi informasi butir yang paling besar yaitu 1,474782 pada butir soal nomor 16, sedangkan fungsi informasi butir yang paling kecil sebesar 0,106319 pada butir soal nomor 9.

Seperti halnya analisis perangkat tes bagian soal pilihan ganda, untuk bagian esai juga dianalisis oleh tiga orang penelaah ahli untuk setiap mata pelajaran. Hasil analisis dapat dipilah menjadi dua kategori butir soal, yaitu butir soal yang diterima tanpa revisi dan diterima dengan revisi. Butir soal dikategorikan diterima tanpa revisi apabila semua kriteria yang terdapat pada

setiap bidang telaah kualitatif yaitu bidang materi, konstruksi dan bahasa telah terpenuhi. Butir soal dikategorikan diterima dengan revisi apabila minimal ada satu kriteria dari ketiga bidang telaah tersebut tidak terpenuhi. Dan apabila seluruh kriteria tidak terpenuhi pada suatu butir soal, berarti butir soal tersebut termasuk butir soal yang jelek atau tidak diterima.

Pada perangkat tes bahasa Indonesia yang memerlukan revisi sebanyak 2 butir soal. Pada perangkat tes Matematika, yang memerlukan perbaikan 2 butir soal, sedangkan pada perangkat tes IPS yang tergolong memerlukan perbaikan atau revisi sebanyak 2 butir soal.

Butir-butir soal perangkat tes Bahasa Indonesia yang direvisi pada bidang materi dan bidang konstruksi tidak ada, sedangkan pada bidang bahasa sebanyak 2 butir soal direvisi. Butir-butir soal perangkat tes Matematika yang direvisi pada bidang materi sebanyak 1 butir soal, sedangkan revisi pada bidang konstruksi tidak ada dan pada bidang bahasa sebanyak 1 butir soal direvisi. Butir-butir soal perangkat tes IPS yang direvisi pada bidang telaah materi tidak ada, yang di revisi pada bidang konstruksi sebanyak 2 butir soal dan yang direvisi pada bidang telaah bahasa tidak ada.

Analisis data statistika untuk perhitungan tingkat kesukaran butir untuk setiap bidang studi, dilakukan dengan cara menjumlah nilai respons semua responden untuk setiap butir pada setiap korektor (Σx_1 dan Σx_2). Selanjutnya nilai Σx_1 dan Σx_2 masing-masing dibagi jumlah responden (633) akan didapatkan mean untuk setiap korektor.

Pada mata pelajaran Bahasa Indonesia, berdasar perhitungan korektor pertama dan kedua yang dikonsultasikan dengan kriteria tingkat kesukaran soal menurut Depdikbud (1994: 104), diperoleh hasil sebagai berikut: Butir-butir soal yang mempunyai tingkat kesukaran rendah adalah butir soal nomor: 39, 40, 41, 42, 45 dan 47. Butir-butir soal yang mempunyai tingkat kesukaran sedang adalah butir soal nomor: 36, 37, 38, 43, 44, 46, 48, 49 dan 50. Butir-butir soal yang mempunyai tingkat kesukaran tinggi tidak ada.

Pada mata pelajaran Matematika, berdasar perhitungan korektor pertama dan kedua yang dikonsultasikan dengan kriteria tingkat kesukaran soal menurut Depdikbud (1994: 104), diperoleh hasil sebagai berikut: Butir-butir soal yang mempunyai tingkat kesukaran rendah adalah butir soal nomor: 31 dan 40. Butir-butir soal yang mempunyai tingkat kesukaran sedang adalah

butir soal nomor: 32, 33, 36, 37, 38 dan 39. Butir-butir soal yang mempunyai tingkat kesukaran tinggi adalah butir soal nomor: 34 dan 35.

Pada mata pelajaran IPS, berdasar perhitungan korektor pertama dan kedua yang dikonsultasikan dengan kriteria tingkat kesukaran soal menurut Depdikbud (1994: 104), diperoleh hasil sebagai berikut: Butir-butir soal yang mempunyai tingkat kesukaran rendah adalah butir soal nomor: 37, 39, 44 dan 45. Butir-butir soal yang mempunyai tingkat kesukaran sedang adalah butir soal nomor: 40, 41, 42, 46 dan 47. Butir-butir soal yang mempunyai tingkat kesukaran tinggi adalah butir soal nomor: 36, 38, 43, 49 dan 50.

Hasil perhitungan daya pembeda soal, dari dua korektor (Rater) terhadap butir soal pelajaran Bahasa Indonesia, dinyatakan: butir tes nomor 36, 37, 38, 43, 44 dan 45 mempunyai indeks diskriminan yang ideal 1, yang berarti butir tersebut semakin mampu membedakan mereka yang menguasai bahan yang diujikan dan mereka yang tidak. Butir tes yang lainnya yaitu nomor 39, 40, 41, 42, 46, 47, 48, 49 dan 50 semuanya mempunyai indeks diskriminasi yang melebihi 0,18, berarti butir tersebut dapat membedakan antara mereka yang menguasai materi pelajaran dan yang tidak menguasai.

Hasil perhitungan daya pembeda soal, dari dua korektor (rater) terhadap butir soal pelajaran Matematika, setelah dikonsultasikan dengan kriteria daya pembeda yang ada pada teori dinyatakan: butir tes nomor 32, 33 merupakan mempunyai indeks diskriminasi ideal 1, yang berarti butir tersebut semakin mampu membedakan mereka yang menguasai bahan yang diujikan dengan yang tidak menguasai bahan yang diujikan. Butir tes nomor 31, 34, 35, 36, 37, 38, 39 dan 40 semuanya termasuk butir tes yang dapat membedakan antara mereka yang menguasai dan yang tidak menguasai materi yang diujikan karena indeks diskriminasi yang dicapai $\geq 0,18$.

Hasil perhitungan daya pembeda soal, dari dua korektor (Rater) terhadap butir soal pelajaran IPS, setelah dikonsultasikan dengan kriteria daya pembeda yang ada dinyatakan: butir tes nomor 40, 41, 42, 46 mempunyai indeks diskriminan yang ideal 1, yang berarti butir tersebut semakin mampu membedakan mereka yang menguasai bahan yang diujikan dan mereka yang tidak. Butir tes yang lainnya yaitu nomor 36, 37, 38, 39, 43, 44, 47, 48, 49 dan 50 semuanya merupakan butir tes yang dapat membedakan *testee* yang pandai dan yang kurang pandai.

Pembahasan

Hasil analisis butir secara teoretis tes objektif terhadap tiga perangkat tes yang dikaji menunjukkan bahwa ketiga perangkat tes tersebut cukup baik. Hal ini ditunjukkan oleh cukup banyak butir-butir soal yang mempunyai kategori **baik tanpa revisi** dan **tidak ada butir soal yang tidak diterima** pada setiap perangkat tes.

Hasil analisis program Rascal yang menghasilkan parameter tingkat kesukaran butir menunjukkan bahwa dari ketiga perangkat tes bahasa Indonesia, Matematika dan IPS hampir seluruh butir soal dapat dikategorikan sebagai alat ukur yang baik. Butir-butir soal yang ditolak atau tidak baik adalah butir soal yang memiliki parameter tingkat kesukaran > 2 atau < -2 dalam skala Logits. Butir-butir soal yang dapat diterima atau baik adalah butir-butir yang memiliki tingkat kesukaran antara ≥ -2 dan ≤ 2 dalam skala Logits pada kurva normal (Hambleton, dkk., 1991: 13).

Butir soal yang cocok dengan model menunjukkan bahwa data mengikuti ICC model logistik 1 parameter yang dipilih. Sebaliknya butir soal yang belum cocok datanya dengan model yang dipilih berarti menunjukkan perbedaan ICC empiris dengan ICC teoretis. Perbedaan tersebut terjadi karena butir soal belum berfungsi sebagaimana mestinya.

Tingkat kemampuan peserta dalam teori respons butir dikenal dengan ciri atau karakteristik peserta. Parameter peserta di dalam model teori respons butir membentuk suatu kontinum. Bentangan kontinum ini tidak dibatasi, kesemuanya bergantung kepada ukuran ciri peserta yang terdapat pada para peserta Ebtanas.

Seperti yang dijelaskan pada landasan teori, secara teoretis nilai baku tingkat kemampuan peserta membentang dari $-\infty$ sampai dengan $+\infty$ dalam skala logits. Namun demikian secara praktis, nilai baku yang berguna hanya terletak antara -4 sampai $+4$ dan adakalanya cukup terletak -3 sampai $+3$.

Hasil analisis data menunjukkan bahwa pada perangkat tes IPS, tingkat kemampuan peserta bervariasi dari $-4,302$ sampai $3,369$. Pada perangkat tes bahasa Indonesia tingkat kemampuan peserta bergerak dari $-2,454$ sampai $2,913$, sedangkan pada perangkat tes Matematika tingkat kemampuan peserta bergerak dari $-2,505$ sampai $2,796$. Setiap peserta memiliki nilai tingkat kemampuan tertentu. Karena jumlah peserta Ebtanas cukup besar maka pada suatu nilai tingkat kemampuan tertentu terdapat lebih

dari satu peserta. Peserta Ebtanas yang memiliki nilai tingkat kemampuan yang sama dalam suatu sampel kelompok tertentu akan memiliki probabilitas yang sama untuk dapat menjawab butir soal Ebtanas dengan benar atau salah.

Nilai fungsi informasi butir pada masing-masing perangkat tes nilainya tidak berbeda jauh, hal ini berarti kualitas ketiga perangkat tes Ebtanas tersebut memiliki tingkat reliabilitas yang hampir sama.

Menurut Hambleton, dkk.(1991), di dalam teori respons butir nilai fungsi informasi butir berbanding terbalik dengan kuadrat dari kekeliruan baku (*standard error*) pada pengestimasiian parameter melalui metode kebolehjadian maksimum. Hal ini berarti jika nilai fungsi informasi suatu butir soal tinggi maka kekeliruan baku akan kecil, sebaliknya jika nilai fungsi informasi suatu butir soal rendah maka kekeliruan baku akan besar.

Pada perangkat tes IPS rata-rata nilai fungsi informasi butir berkisar pada nilai 0,1..... tetapi pada butir 16 nilai fungsi informasi butir maksimumnya jauh lebih besar yaitu 1,474782 dan ini merupakan butir soal yang mempunyai nilai fungsi informasi tertinggi dibanding nomor-nomor butir lainnya pada ketiga perangkat tes Ebtanas. Butir soal yang demikian berarti butir ini memiliki kekeliruan baku (*standard error of estimation*) yang kecil sekali.

Hasil analisis secara kualitatif terhadap tiga perangkat tes yang dikaji menunjukkan bahwa ketiga perangkat tes tersebut cukup baik. Hal ini ditunjukkan oleh cukup banyak butir-butir soal yang baik tanpa revisi dan ada beberapa butir soal yang perlu direvisi. Perangkat tes IPS sebanyak 13 butir yang diterima (86,67%). Perangkat tes Matematika sebanyak 8 butir yang diterima (80%) dan perangkat tes bahasa Indonesia yang diterima sebanyak 13 butir (86,67%).

Tingkat kesukaran butir pada ketiga perangkat tes Ebtanas SD se Kabupaten Jepara ditentukan berdasar hasil penyekoran dari setiap Rater. Dalam hal ini rater-nya ada 2 orang yaitu guru-guru SD yang ada di Kabupaten Jepara.

Tingkat kesukaran butir antara rater pertama dan kedua hampir sama dan jika ada selisih, maka selisihnya kecil sekali. Depdikbud (1994: 104) merinci indeks kesukaran butir sebagai berikut: $P = 0,00$ sampai dengan $0,30$ menunjukkan butir yang sukar. $P = 0,31$ sampai dengan $0,70$ menunjukkan

butir yang sedang dan $P = 0,71$ sampai dengan $1,00$ menunjukkan butir yang mudah

Berdasarkan hasil analisis tingkat kesukaran butir pada perangkat tes IPS, tingkat kesukaran butir pada rater pertama mulai dari $0,125$ sampai $0,91$, sedangkan pada rater kedua mulai dari $0,13$ sampai $0,915$. Pada perangkat tes IPS ini, yang memiliki tingkat kesukaran sedang atau baik adalah sebanyak 6 butir (40%) sedangkan sisanya termasuk butir yang tidak baik, terdiri dari 4 butir (26,67%) termasuk butir yang mudah dan 5 butir soal (33,33%) termasuk butir soal yang sukar.

Daya beda butir pada ketiga perangkat tes Ebtanas SD di Kabupaten Jepara ditentukan berdasarkan hasil penyekoran dari rater. Dalam hal ini raternya terdiri atas 2 orang yaitu guru-guru SD yang ada di Kabupaten Jepara. Besarnya penyekoran antar-rater pertama dan rater kedua tidak terlalu beda sehingga menghasilkan indeks daya beda yang hampir sama juga.

Ketentuan besarnya daya beda menurut teori adalah untuk ukuran sampel > 350 , jika indeks daya beda hasil perhitungan $\geq 0,18$ maka indeks daya beda tersebut termasuk indeks daya beda yang baik (Depdikbud, 1994: 106), sedangkan menurut Thorndike (1991) jika butir yang memiliki indeks daya beda kurang dari $0,2$ dapat langsung dibuang.

Untuk menghitung konsistensi penyekoran yang dilakukan oleh dua orang rater dalam melakukan penyekoran terhadap perangkat tes uraian digunakan perhitungan reliabilitas antar-rater. Melalui perhitungan diketahui semua butir pada setiap perangkat tes terdapat konsistensi penyekoran antara dua orang rater, dimana indeks reliabilitasnya $> 0,7$.

Penutup

Berdasarkan temuan hasil penelitian dan pembahasan dapat dikemukakan beberapa kesimpulan analisis sebagai berikut:

- Hasil analisis kualitatif tes objektif pada ketiga perangkat tes yang dianalisis secara umum memiliki kualitas yang cukup baik.
- Hasil estimasi parameter butir yang berupa tingkat kesukaran butir tes objektif pada ketiga perangkat tes secara umum memiliki kualitas yang cukup baik.
- Butir-butir soal tes objektif pada ketiga perangkat tes rata-rata mempunyai persentase yang besar untuk kecocokan dengan model *Rasch*.

- Tingkat kemampuan peserta Ebtanas tergolong sedang.
- Nilai fungsi informasi butir pada masing-masing perangkat tes nilainya tidak berbeda jauh, berarti kualitas ketiga perangkat tes Ebtanas memiliki tingkat reliabilitas yang hampir sama.
- Hasil analisis kualitatif tes esai pada ketiga perangkat tes Ebtanas SD umumnya sebagian besar mempunyai butir soal yang baik.
- Hasil analisis empiris tentang tingkat kesukaran butir pada tes esai untuk perangkat tes Bahasa Indonesia diperoleh butir yang baik adalah 9 butir soal (60%). Pada perangkat tes Matematika yang terdiri dari 10 butir soal diperoleh butir yang baik adalah 6 butir soal (60%), sedangkan pada perangkat tes IPS diperoleh butir soal yang baik adalah 6 butir soal (40%).
- Daya pembeda pada semua perangkat tes menggambarkan butir-butir soalnya dapat membedakan antara mereka yang menguasai materi dengan yang tidak menguasai materi.
- Reliabilitas hasil penyekoran (konsistensi penyekoran) pada semua perangkat tes, diketahui semua hasil penyekoran konsisten antara korektor pertama maupun kedua kecuali pada perangkat tes IPS yaitu butir No.47 yang tidak konsisten.

Daftar Pustaka

- Azwar, Saifuddin. (1996). *Tes prestasi, fungsi dan pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- . (1997). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Depdikbud. (1994). *Pedoman penulisan soal*. Jakarta: Ditjen Dikdasmen.
- . (1995). *Bank soal: Buletin pengujian dan penilaian*. Jakarta: Puslitbang-sisjian. Balitbang Depdikbud.
- Ebel, Robert L. (1972). *Essentials of educational measurement*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Hambleton, Ronald K. dan Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer. Nijhoff Publishing.

Karakteristik Soal Tes Ebtanas Sekolah Dasar di Kabupaten Jepara

- Hambleton, Ronald K., Swaminathan, H., dan Rogers H.Jane. (1991). *Fundamentals of item response theory*. Newbury Park, California: Sage Publications.
- Hulin, Charles L., Drasgow, Fritz dan Parsons, Charles K. (1983). *Item response theory: Application to psychological measurement*. Homewood, Illinois: Dow Jones-Irwin.
- Lord, Frederic M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
- Mardapi, J. (1991). Konsep Dasar Teori Respons Butir : Perkembangan dalam Bidang Pengukuran Pendidikan. IKIP Yogyakarta: *Cakrawala Pendidikan* Nomor 3, Tahun X, November 1991.
- . (1994). *Analisis butir dengan teori tes klasik dan teori respons butir*. IKIP Yogyakarta: Laporan Penelitian.
- . (1997). *Ragam bentuk evaluasi*. UGM: Makalah Semiloka Evaluasi Sistem Penilaian dan Pengukuran Prestasi Belajar Mahasiswa UGM Tanggal 13 dan 14 Mei 1997.
- Mehrens, William A. dan Lehmann, Irvin J. (1973). *Measurement and evaluation in education and psychology*. New York: Hoit, Rinehart & Winston, Inc.
- Naga, Dali S. (1992). *Pengantar teori sekor: Pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Ratna Sayekti Rusli. (1988). *Tes dan pengukuran dalam pendidikan*. Jakarta: Depdikbud Ditjen Dikti, Proyek Pengembangan LPTK.
- Suryabrata, Sumadi. (1984). *Pembimbing ke Psikodiagnostik, Edisi II Cetakan 1*. Yogyakarta: Rake Press.
- . (1987). *Pengembangan tes hasil belajar*. Jakarta: Rajawali.
- Thorndike, R.L. dan Hagen, E.P. (1977). *Measurement and evaluation in psychology and education*. 4th ed. New York: John Wiley and Son.
- Umar, J. (1991). *Pengantar penilaian pendidikan (makalah)*. Jakarta: PUSISJIAN, Balitbang, Depdikbud.
- Wright, B.D. dan Stone, M.H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.