

KUALITAS TES UJIAN AKHIR SEKOLAH BERSTANDAR NASIONAL (UASBN) IPA SD TAHUN PELAJARAN 2007/2008 DI KOTA KENDARI

Oleh:
Muh. Nurung

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas tes ujian akhir sekolah berstandar nasional (UASBN) yang digunakan di kota Kendari pada tahun pelajaran 2007/2008, untuk mata ujian IPA.

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan survei. Populasi penelitian meliputi paket tes dan respons siswa terhadap butir soal IPA UASBN di kota Kendari. Penentuan sampel paket tes menggunakan teknik *purposive sampling*. Paket yang dipilih adalah paket utama P1 yang terdiri dari 40 butir soal pilihan ganda dan selanjutnya ditelaah secara kualitatif (teoretis). Sampel respon siswa ditentukan dengan menggunakan teknik *stratified proportional random sampling*. Respons siswa sebanyak 600 sampel, dianalisis secara kuantitatif (empiris) berdasarkan teori tes klasik program *ITEMAN* versi 3.0 dan teori respons butir program *BIGSTEPS* versi 2.3 model Rasch.

Hasil penelitian menunjukkan bahwa: (1) Kualitas tes secara kualitatif (teoretis) kurang baik, terdapat 27 butir soal (67,5%) sesuai dengan kriteria penelaahan, sedangkan sisanya 13 butir soal (32,5%) tidak sesuai dengan kriteria; (2) Secara kuantitatif (empiris) berdasarkan teori tes klasik dengan program *ITEMAN* diperoleh indeks keandalan tes 0,826. Ditemukan sebanyak 24 butir soal (60%) kategori baik dan 16 butir soal (40%) kategori tidak baik sehingga kualitas tes menurut teori tes klasik dinyatakan kurang baik; (3) Secara kuantitatif (empiris) berdasarkan teori respons butir dengan program *BIGSTEPS* diperoleh fungsi informasi tes 0,838 yang berarti handal. Ditemukan 35 butir soal (87,5%) kategori baik dan 5 butir soal (12,5%) kategori tidak baik sehingga kualitas tes menurut teori respon butir dinyatakan baik; dan (4) Indeks konsistensi hasil analisis butir soal menurut teori tes klasik dan teori respon butir sebesar 0,462 (sedang) yang berarti hasil analisis menggunakan kedua metode tersebut cukup konsisten. Keseluruhan jumlah butir soal yang baik menurut ketiga metode analisis yang digunakan adalah 19 butir (47,5%), sedangkan butir soal yang tidak baik sebanyak 21 butir (52,5%).

Kata Kunci: *Kualitas tes, UASBN*

ABSTRACT

MUH. NURUNG: *The Quality of the Final Examination Test of IPA SD of National Standard School in the Academic Year of 2007/2008 in Kendari City of South East Sulawesi.* Thesis. Yogyakarta: Graduate School, State University of Yogyakarta, 2008.

This research aims at finding out the quality of the final examination test of IPA SD (Natural Science for Elementary School) of National Standard School in the academic year of 2007/2008 in Kendari City, South East Sulawesi.

This research employed the survey approach. The research population consisted of test packages and student's response toward the IPA subject test items of UASBN in Kendari City. The test package sample was established by means of purposive sampling technique. The package chosen was a P1 main package that consists of 40 multiplechoice items and then analyzed qualitatively. The student's response sample was determined by stratified proportional random sampling technique. The number of student's response was 600, which were qualitatively analyzed on the basis of a classical test theory making use of ITEMAN program of 3.0 version and item responses theory using BIGSTEPS program of 2.3 Rasch model.

The results of the research show the following. (1) The quality of the test is *not quite good*, there are 27 items (67.5%) meeting the analysis criterion, while 13 test items (32.5%) do not; (2) Based on a classical test theory using ITEMAN program it is found that the test reliability index is 0.826, there are 24 test items (60%) in *good* category and 16 test items (40%) are *not in good* category so that the overall test quality is *not quite good*; (3) Based on items responses theory using the BIGSTEPS program it is found that the test information function is 0.838 which means the test is reliable. There are 35 test items (87.5%) in *good* category and 5 items (12.5%) is *not good* category to make the overall test quality falls into *good* category; (4) the consistency index of item analysis result based on classical test theory and item responses theory is 0.462 which is in *medium* category, meaning the results of the analysis using both methods are consistent. The total number of good test items based on the three of analysis methods of analysis is 19 (47.5%), while the bad test items are 21 (52.5%).

Key word : *Test quality, UASBN*

Pendahuluan

Evaluasi merupakan salah satu rangkaian kegiatan dalam meningkatkan kualitas, kinerja atau produktivitas suatu lembaga dalam melaksanakan programnya Djemari Mardapi (2008: 8). Oleh karena itu, evaluasi merupakan salah satu subsistem yang penting dalam sistem pendidikan. Dalam Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, evaluasi diatur dalam Bab XVI Pasal 57, 58, dan 59. Pelaksanaan evaluasi bertujuan untuk mengukur dan mengendalikan mutu pendidikan.

Penjabaran lebih lanjut tentang pelaksanaan evaluasi dinyatakan dalam Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan. Pada pasal 63 ayat (1) menyebutkan bahwa penilain pendidikan pada jenjang pendidikan dasar dan menengah terdiri atas: (a) penilaian hasil belajar oleh pendidik, (b) penilaian hasil belajar oleh satuan pendidikan, dan (c) penilaian hasil belajar oleh pemerintah. Penilaian hasil belajar oleh pendidik dan satuan pendidikan merupakan bentuk evaluasi internal (*internal evaluation*). Berkenaan dengan evaluasi oleh pemerintah, sifatnya sebagai evaluasi eksternal (*external evaluation*) yang sasarannya adalah peserta didik, termasuk di dalamnya hasil belajar peserta didik.

Penilaian hasil belajar peserta didik pada jenjang Sekolah Dasar (SD), pada tahun pelajaran 2007/2008 berdasarkan Keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor 39 Tahun 2007 ditetapkan sistem penilaiannya dalam bentuk Ujian Akhir Sekolah Berstandar Nasional (UASBN). Khusus penyelenggaraan Ujian Akhir Sekolah Berstandar Nasional (UASBN) tahun pelajaran

2007/2008, komposisi butir soal, sebanyak 25% ditetapkan oleh Badan Standar Nasional Pendidikan (BSNP) dan berlaku secara nasional, serta 75% soal ditetapkan oleh penyelenggara UASBN tingkat provinsi (Peraturan Mendiknas, 2007b).

Ujian Akhir Sekolah Berstandar Nasional (UASBN) jenjang SD dalam pelaksanaannya berpedoman pada Keputusan Badan Standar Nasional Pendidikan Nomor 983/BSNP/XI/2007 tentang Prosedur Operasi Standar (POS) Ujian Akhir Sekolah Berstandar Nasional untuk Sekolah Dasar/Madrasah Ibtidaiyah/Sekolah Dasar Luar Biasa (SD/MI/SDLB) Tahun Pelajaran 2007/2008. Bentuk tes yang digunakan diantaranya berupa tes tertulis (*paper and pencil test*). Tes tertulis merupakan teknik penilaian yang seringkali digunakan untuk menilai prestasi belajar siswa. Melalui tes prestasi belajar, dapat diperoleh informasi yang dapat menggambarkan kemampuan siswa (Bauer,2000 :4). Oleh karena itu, pengelolaan ujian dan mutu bahan ujian yang digunakan perlu mendapat perhatian agar hasil tes dapat mencerminkan kemampuan siswa yang sebenarnya.

Secara sederhana Allen & Yen (1979: 1) menyebut tes sebagai “*a test is device for obtaining a sample of an individual’s behavior*”. Hal yang hampir sama juga dikemukakan oleh Gronlund (1985: 5) yang mendefinisikan tes sebagai instrumen atau prosedur sistematis untuk mengukur perilaku sampel. Ahli pengukuran yang lain, Djaali (2006: 57) menyatakan tes adalah suatu cara atau alat untuk mengadakan penilaian yang berbentuk suatu tugas atau serangkain tugas yang harus dikerjakan oleh siswa atau sekelompok siswa sehingga menghasilkan nilai tentang tingkah laku atau prestasi siswa sebagai peserta didik.

Klasifikasi tes dikemukakan oleh beberapa pakar, diantaranya (Gronlund, 1982: 19 & Santoso Murwani, 2006: 18-20) yang membedakan tes menjadi empat yaitu: 1) tes penempatan, 2) tes formatif, 3) tes diagnostik, dan 4) tes sumatif. Tes juga dapat dibedakan menjadi *achievement test* atau tes prestasi belajar dan *learning outcome test* atau tes hasil belajar (Tim Puslitbang Sisjian, 1999: 15). Dalam tes prestasi belajar, yang hendak diukur ialah tingkat kemampuan seorang siswa dalam menguasai bahan pelajaran yang diajarkan kepadanya. Oleh karenanya, kedudukan tes prestasi dalam pengambilan keputusan sangat penting. Meski demikian, menurut Bauer (2000 :4), skor yang diperoleh siswa dalam tes prestasi belajar kemungkinan tidak sepenuhnya mencerminkan kualitas pembelajaran di sekolah. Hal ini dikarenakan kemampuan siswa tidak semata dipengaruhi oleh pembelajaran di sekolah, tetapi oleh banyak faktor.

Bentuk soal dalam tes prestasi belajar, secara umum dapat dikelompokkan menjadi dua kategori yaitu: 1) tes uraian, terdiri dari uraian bebas, uraian terbatas atau isian singkat, uraian berstruktur, dan 2) tes objektif, terdiri dari pilihan benar-salah, pilihan ganda, dan menjodohkan. Didalam penelitian ini, digunakan tes bentuk pilihan ganda. Bentuk soal ini, jawabannya harus dipilih dari beberapa kemungkinan jawaban yang telah disediakan. Penggunaan tes pilihan ganda, pada umumnya dijumpai pada ujian yang berskala besar/massal karena sifatnya yang obyektif dan mudah penskorannya. Bentuk soal ini juga dianggap pilihan yang tepat untuk ujian akhir dimana bahan pelajaran yang hendak diujikan biasanya cukup banyak. Dilihat dari strukturnya, bentuk soal pilihan ganda terdiri dari pokok soal (*stem*) dan pilihan

jawaban (*option*). Pilihan jawaban terdiri atas satu kunci jawaban dan yang lainnya pengecoh (*distraktor*). Pokok soal (*stem*) dapat berupa pertanyaan atau pernyataan tidak lengkap sebagaimana dinyatakan Thorndike (2005: 448) berikut ini:

the multiple-choice item consists of two parts: the stem, which presents the problem, and the list of possible answer or options. In the standard form of the item, one of the options is the correct or best answer and the others are foils or distractor. The stem of the item may be presented either as a question or as an incomplete statement.

Kualitas tes, termasuk bentuk tes pilihan ganda (*dikotomi*) dapat diungkap melalui analisis butir soal secara teoretis (*telaah*) dan analisis empiris. Analisis butir soal secara kualitatif dilakukan untuk menilai butir soal ditinjau dari aspek materi, konstruksi, dan bahasa. Analisis secara kuantitatif menekankan pada analisis karakteristik butir soal secara empiris. Karakteristik butir soal antara lain meliputi indeks kesukaran (*p*), daya beda (*d*), dan distribusi respons.

Analisis secara empiris dapat menggunakan pendekatan tes klasik maupun pendekatan tes modern (*IRT*). Menurut Djemari Mardapi (2008: 32), pada teori tes klasik, besarnya skor tampak (*X*) peserta ujian merupakan penjumlahan dari skor murni (*T*) dan skor kesalahan pengukuran (*E*). Secara matematis, dapat dituliskan dalam bentuk formula berikut ini.

$$X = T + E \quad (1)$$

Teori tes klasik memiliki beberapa asumsi seperti dinyatakan Schumacker (2005: 1) yaitu: (1) tidak ada korelasi antara skor yang sebenarnya dan skor kesalahan, (2) rerata kesalahan acak pengukuran sama dengan nol, dan (3) skor

kesalahan pada test paralel tidak berkorelasi. Dengan dasar asumsi tersebut, maka dikembangkan sejumlah formula untuk mengestimasi indeks keandalan, indeks kesahihan tes, dan indeks kesukaran serta daya beda dari suatu tes. Selanjutnya, sejumlah ahli mengembangkan program komputer untuk mengestimasi parameter butir berdasarkan teori tes klasik. Pada penelitian ini, digunakan program *ITEMAN MicroCAT (tm) Testing System, versi 3.00*.

Sayangnya, seperti diungkapkan Saifuddin Azwar (2005: 79) bahwa teori tes klasik memiliki keterbatasan yang mendasar antara lain pertama, hasil estimasi parameter tergantung pada karakteristik peserta ujian (*group dependent*). Hal ini berimplikasi pada tingkat kesukaran soal akan menjadi rendah jika tes diujikan pada kelompok peserta tes berkemampuan tinggi, dan sebaliknya jika tes diujikan pada peserta dengan kemampuan rendah, maka tingkat kesukaran tes itu akan tinggi. Kedua, hasil estimasi kemampuan peserta tergantung pada karakteristik butir soal (*item dependent*).

Mengatasi kelemahan pada teori tes klasik, maka para ahli pengukuran mengembangkan model yang tidak terikat dengan sampel (*sample free*). Model ini selanjutnya dikenal tes modern atau teori respons butir. Menurut teori respon butir, perilaku seseorang dapat dijelaskan oleh oleh karakteristik orang yang bersangkutan sampai pada batas-batas tertentu (Djemari Mardapi, 2008: 145).

Hambleton & Swaminathan (Sinharay & Almond, 2007: 245) menyatakan bahwa teori respons butir (IRT) merupakan salah satu cara untuk menilai kelayakan butir dengan membandingkan rerata penampilan butir terhadap tampilan bukti

kemampuan kelompok yang diramalkan oleh model. Tujuan utama teori respons butir dikembangkan adalah untuk mengatasi kelemahan teori tes klasik yang tidak *independent* terhadap kelompok peserta yang mengerjakan tes maupun terhadap tes yang diujikan.

Bagian penting dari teori respons butir adalah probabilitas jawaban benar peserta tes, parameter butir dan parameter peserta tes dihubungkan melalui suatu fungsi matematik atau model formula matematik. Dalam formula ini, nilai kemungkinan peserta tes menjawab soal dipahami sebagai fungsi logistik perbedaan parameter yang dimasukkan ke dalam model. Hal ini sebagaimana dijelaskan oleh Hulin, Drasgow, dan Parsons (1983: 14), dalam pernyataan berikut.

an item response theory includes a set of propositions concerned with individuals' responses to items used for psychological measurement. An essential part of each IRT presented in this book is a mathematical function that relates the probability of some type of response to an item by an individual to certain characteristics of the individual and the item.

Terdapat beberapa model pengukuran yang termasuk kategori psikometrika modern atau apa yang dikenal dengan sebutan *latent trait model* atau *Item Response Theory (IRT)*. Model pengukuran tersebut dibedakan berdasarkan jumlah parameter butir yang dimasukkan ke dalam model, yaitu model satu parameter (1P) atau model Rasch, dua parameter (2P), dan tiga parameter (3P). Dalam penelitian ini, akan digunakan model satu parameter (model Rasch) dengan bantuan program *BIGSTEPS* versi 2.30. Pemilihan model Rasch didasarkan pada pernyataan Ivailo Partchev

(2004: 10) yang menyebutkan bahwa untuk butir soal dikotomi dapat menggunakan model satu parameter (1P).

Menurut Hambleton, Swaminathan, & Rogers (1991: 12-13) persamaan matematik fungsi karakteristik butir model Rasch dapat dinyatakan seperti berikut ini.

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

Keterangan:

$P_i(\theta)$ = peluang menjawab benar peserta yang berkemampuan θ pada butir ke-i

b_i = tingkat kesukaran butir ke-i

e = bilangan transendental yang bernilai 2,718

n = banyaknya butir tes

D = konstanta bernilai 1,7 sebagai simpangan baku distribusi logistik

Persamaan tersebut di atas, menunjukkan bahwa lengkungan kurva model Rasch hanya ditentukan oleh parameter tingkat kesukaran butir b_i dan parameter ciri peserta θ yang menjawab butir tersebut. Jadi peluang seseorang menjawab benar suatu butir merupakan fungsi dari kemampuan peserta dan tingkat kesukaran butir. Tingkat kesukaran butir soal bergerak dari $-\infty$ sampai $+\infty$. Meski demikian nilai yang berarti biasanya bergerak dari skala -3 sampai +3 dalam satuan *logit* (*log odd unit*). Butir yang memiliki tingkat kesukaran (b_i) dekat atau terletak di bawah skala -2,00 menunjukkan bahwa butir tes tersebut termasuk kategori mudah. Butir yang memiliki tingkat kesukaran (b_i) dekat atau terletak di atas skala +2,00 menunjukkan butir tes tersebut termasuk kategori sukar. Butir yang dinyatakan baik adalah butir yang

memiliki tingkat kesukaran (b_i) berkisar $-2 \leq b_i \leq +2$ (Hambleton, Swaminathan, & Rogers, 1991: 13).

Pada teori respons butir, dikenal fungsi informasi yang digunakan untuk menggambarkan sumbangan/kekuatan butir soal dalam mengungkap *latent trait*. Fungsi informasi pada model logistik satu parameter (1P) memenuhi persamaan berikut.

$$I_i(\theta) = \frac{D^2 e^{D(\theta-b_i)}}{[1 + e^{D(\theta-b_i)}]^2} \quad (3)$$

(Naga, 1992: 319)

Keterangan:

$I_i(\theta)$ = Fungsi informasi butir

D = Konstanta, yang nilainya 1,7

θ_i = Skala kemampuan

b_i = Tingkat kesukaran butir i

Berdasarkan persamaan fungsi informasi butir model logistik satu parameter (1 P) maka, nilai fungsi informasi akan maksimum pada saat $\theta = b_i$ atau ketika $I_i(\theta) = 0,7225$. Fungsi informasi tes merupakan jumlah dari fungsi informasi butir (Reid *et al*, 2007: 189). Hal ini berarti, fungsi informasi tes akan tinggi jika butir soal penyusunnya mempunyai fungsi informasi yang tinggi pula. Fungsi informasi tes secara matematis dapat dituliskan sebagai berikut.

$$I(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_{i_c}(\theta)Q_i(\theta)} \quad (4)$$

(Hambleton dan Swaminathan, 1985: 104)

Parameter butir dan parameter peserta merupakan hasil estimasi, sehingga kebenarannya bersifat probabilitas dan tidak bebas terhadap kesalahan pengukuran. Untuk menentukan nilai kesalahan pengukuran skor seorang peserta tes maka diperlukan informasi yang berkenaan dengan *Standard Error of Measurement (SEM)*.

Menurut teori respons butir, *Standard Error of Measurement (SEM)* berkaitan erat dengan fungsi informasi. Fungsi informasi dengan *SEM* mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka *SEM* semakin kecil dan sebaliknya maka bentuk hubungan keduanya memenuhi persamaan:

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \quad (5)$$

(Hambleton dan Swaminathan, 1985: 104)

Berdasarkan kajian teori tersebut di atas, dapat ditentukan karakteristik butir soal. Menurut teori tes klasik, besarnya tingkat kesukaran (p) adalah 0,30 sampai 0,70 (Allen & Yen, 1974: 121 & Sumarna Surapranata, 2006: 47). Daya pembeda (r_{bis}) yang baik minimum 0,30 (Ebel, 1979 : 267). Pengecoh dikatakan berfungsi dengan baik jika paling sedikit dipilih oleh 5% peserta tes, dan indeks kehandalan tes yang baik minimum 0,7 (Kaplan & Saccuso, 1982: 106). Adapun ketentuan yang digunakan untuk menyatakan butir soal yang baik dan tidak baik menurut teori respons butir dengan menggunakan program *BIGSTEP* atau pendekatan model Rasch adalah apabila butir cocok dengan model, dan memiliki tingkat kesukaran berkisar $-2,0 \leq b_i \leq +2,0$.

Berdasarkan kajian tersebut di atas, penelitian terhadap kualitas tes Ujian Akhir Sekolah Berstandar Nasional (UASBN) di provinsi Sulawesi Tenggara, khususnya di kota Kendari, bertujuan untuk: (1) Mendeskripsikan kualitas tes UASBN mata pelajaran IPA SD tahun pelajaran 2007/2008 di kota Kendari ditinjau dari aspek teoretis; (2) Mendeskripsikan kualitas tes UASBN mata pelajaran IPA SD tahun pelajaran 2007/20078 di kota Kendari ditinjau dari aspek empiris; (3) Mendeskripsikan seberapa besar tingkat konsistensi hasil analisis butir soal antara teori tes klasik dengan teori respons butir model Rasch.

Metode Penelitian

Penelitian ini merupakan penelitian deskriptif kuantitatif dengan pendekatan *surve*. Pelaksanaan penelitian dimulai pada bulan Maret sampai dengan Juni 2008 di kota Kendari provinsi Sulawesi Tenggara. Populasi dalam penelitian ini adalah paket tes dan respons siswa peserta UASBN mata pelajaran IPA tahun pelajaran 2007/2008 di kota Kendari sebanyak 4796 siswa dari 117 sekolah. Pengambilan sampel penelitian berupa paket tes utama P1 ditentukan dengan teknik *purposive sampling* sedangkan sampel respons siswa peserta UASBN mata pelajaran IPA tahun pelajaran 2007/2008 sebanyak 600 lembar jawaban siswa ditentukan dengan teknik *stratified proportional random sampling*.

Teknik pengumpulan data dalam penelitian ini berupa: (1) Daftar cek (kartu telaah), digunakan untuk melakukan telaah butir soal guna memperoleh data tentang butir soal yang memenuhi kriteria dilihat dari aspek materi, konstruksi, dan bahasa. Instrumen yang digunakan berupa kartu telaah butir soal yang ditetapkan oleh Pusat

Penilaian Pendidikan (Puspendik) Departemen Pendidikan Nasional. (2) Dokumen, digunakan untuk memperoleh data dalam penelitian ini antara lain: Standar Kompetensi Lulusan UASBN SD mata pelajaran IPA, spesifikasi soal UASBN SD mata pelajaran IPA, lembar soal UASBN SD mata pelajaran IPA, kunci Jawaban UASBN SD mata pelajaran IPA, dan lembar jawaban siswa peserta UASBN SD mata pelajaran IPA Tahun Pelajaran 2007/2008 di kota Kendari.

Data yang diperoleh dalam penelitian ini, selanjutnya dianalisis secara kualitatif (teoretis) dan secara kuantitatif (empiris). Analisis secara kualitatif dilakukan dalam dua (2) tahap yakni pertama, telaah awal perangkat tes dimaksudkan untuk menyeleksi butir soal yang harus dianulir (*drop*) seperti soal yang tidak ada kuncinya, memiliki kunci ganda atau memiliki konstruksi soal yang tidak lengkap. Kedua, telaah butir soal (*item review*) berdasarkan pertimbangan professional (*expert judgment*). Tahapan ini dimaksudkan untuk melihat perilaku soal yang diharapkan, ditinjau dari aspek materi, konstruksi, dan bahasa.

Analisis empiris berdasarkan pendekatan teori tes klasik dilakukan dengan bantuan program *Item and Test Analysis (ITEMAN)* Versi 3.00. Analisis ini akan menghasilkan karakteristik butir soal dan perangkat tes berupa statistik. Statistik butir tes, meliputi: (1) tingkat kesukaran, (2) daya beda, dan (3) efektivitas distraktor. Statistik perangkat tes, antara lain: mean, median, indeks keandalan, kemencengan, dan kesalahan baku pengukuran. Kualitas perangkat tes yang baik secara empiris berdasarkan teori tes klasik, apabila tingkat kesukaran terletak pada interval : $0,30 \leq$

$p \leq 0,70$, korelasi *biserial* (r_{bis}) butir $\geq 0,30$, korelasi biserial pilihan jawaban semua bernilai negatif selain kunci dan pilihan jawaban dipilih oleh minimal 5% peserta tes.

Analisis butir soal dengan pendekatan teori respons butir, dalam penelitian ini digunakan program *MicroCat BIGSTEPS* versi 2.30 atau lebih dikenal dengan model Rasch. Berdasarkan hasil (*out put*) dari analisis menggunakan software ini, akan diperoleh informasi yang berkaitan dengan tingkat kesukaran butir soal (*measure difficulty*) dalam bentuk skala *LOGIT*, indeks kesalahan pengukuran (*standard error of measurement*), kecocokan antara data dan model (*INFIT dan OUTFIT*), korelasi *point biserial*, serta estimasi kemampuan peserta.

Parameter tingkat kesukaran butir yang terletak pada skala logits $b_i < -2,00$ dikategorikan soal yang mudah, sedangkan butir soal dengan $b_i > 2,00$ dikategorikan soal yang sukar. Butir soal yang tidak cocok dengan model menunjukkan bahwa butir tersebut tidak berperilaku secara konsisten sebagaimana yang diharapkan oleh model. Adapun ketentuan yang digunakan untuk menyatakan butir soal yang baik, apabila butir cocok dengan model, dan memiliki tingkat kesukaran berkisar $-2,0 \leq b_i \leq +2,0$.

Selanjutnya, kualitas tes secara kualitatif dan kuantitatif ditentukan berdasarkan persentase butir yang tidak baik. Menurut Kusmiyati (2005), kualitas tes dapat ditentukan berdasarkan kriteria : 0% s.d 10% sangat baik; 11% s.d 20% baik; 21% s.d 30% cukup baik; 31% s.d 40% kurang baik; lebih dari 40% tidak baik.

Penelitian ini, menggunakan pendekatan teori tes klasik maupun teori tes respon butir dalam menganalisis butir soal secara kuantitatif. Kedua pendekatan

mempunyai asumsi dengan persyaratan yang berbeda. Oleh karena itu, konsistensi setiap butir soal terhadap kedua metode dianalisis menggunakan tabel *phi* (Φ). Selanjutnya, hasil analisis konsistensi dapat diinterpretasikan pada nilai korelasi : $0,00 \leq \Phi \leq 0,20$ korelasi sangat rendah; $0,21 \leq \Phi \leq 0,40$ korelasi rendah; $0,41 \leq \Phi \leq 0,60$ korelasi sedang; $0,61 \leq \Phi \leq 0,80$ korelasi tinggi; $0,81 \leq \Phi \leq 1,00$ korelasi sangat tinggi.

Hasil-hasil Pembahasan

Telaah terhadap butir soal pilihan ganda IPA UASBN SD di kota Kendari, dilakukan oleh 6 (enam) orang penelaah yang memiliki kompetensi di bidang materi, konstruksi, dan bahasa. Hasilnya menunjukkan bahwa butir soal IPA UASBN SD tahun pelajaran 2007/2008 di kota Kendari secara kualitatif (teoretis) masih kurang baik. Hal ini didasarkan pada hasil analisis terhadap 40 butir soal pilihan ganda dengan 17 kriteria penelaahan. Sebanyak 13 butir soal (32,5%) yang tidak memenuhi kriteria, dengan rincian butir soal: 1, 3, 4, 5, 9, 11, 12, 20, 23, 25, 26, 27, dan 36.

Analisis butir soal secara empiris dengan pendekatan teori tes klasik program *ITEMAN* meliputi tiga aspek yaitu tingkat kesukaran, daya beda, dan distribusi jawaban. Berdasarkan *output* program *ITEMAN* diperoleh data soal yang baik sebanyak 24 butir (60%) yakni nomor 2, 4, 6, 7, 9, 10, 11, 13, 14, 15, 16, 17, 19, 22, 23, 24, 29, 31, 32, 33, 34, 37, 38, 40. Soal yang tidak baik sebanyak 16 butir (40%) yaitu nomor 1, 3, 5, 8, 12, 18, 20, 21, 25, 26, 27, 28, 30, 35, 36, 39. Indeks keandalan tes sebesar 0,828 (handal) dan indeks kesalahan baku pengukuran (*SEM*) 2,658.

Berdasarkan hasil analisis dengan program *BIGSTEPS* pada tabel di atas, diperoleh informasi bahwa butir soal yang cocok dengan model (*fit*) sebanyak 37 butir (92,5%) yakni butir nomor 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39 dan 40. Sedangkan butir soal yang tidak cocok dengan model sebanyak 3 butir (7,5%) yakni butir nomor 1, 36 dan 39. Nilai maksimum fungsi informasi butir soal terletak pada kemampuan peserta 0 (nol) dengan nilai informasi 17,06 serta kesalahan pengukuran (*SEM*) 0,242. Jumlah butir soal yang memiliki nilai maksimum $\geq 0,7$ sebanyak 83,8% yang menunjukkan tes ini handal.

Nilai konsistensi butir soal dinyatakan dengan indeks konsistensi (Φ). Hasil analisis terhadap naskah tes IPA UASBN SD tahun pelajaran 2007/2008 di kota Kendari menunjukkan bahwa indeks konsistensi (Φ) butir soal adalah 0,462 artinya bahwa tes memiliki konsistensi yang sedang (cukup baik) terhadap kedua teori tes yang digunakan yakni teori tes klasik dan teori respons butir.

Kesimpulan

Mencermati hasil analisis data penelitian dan pembahasannya pada bab sebelumnya tentang analisis butir soal secara teoretis maupun empiris, maka dapat dibuat kesimpulan sebagai berikut:

1. Kualitas tes UASBN IPA SD tahun pelajaran 2007/2008 di kota Kendari berdasarkan analisis kualitatif (teoretis) dikategorikan kurang baik. Hal ini tergambar dari banyaknya butir soal yang tidak memenuhi kriteria yaitu sebanyak

13 butir (32,5%) dengan rincian butir soal nomor 1, 3, 4, 5, 9, 11, 12, 20, 23, 25, 27, 34, dan 36.

2. Analisis kuantitatif (empiris) butir soal dengan pendekatan teori tes klasik menggunakan program *ITEMAN* diperoleh informasi indeks keandalan tes sebesar 0,826 dengan *SEM* 2,658 yang berarti tes andal dalam melakukan fungsi ukurnya. Sebanyak 16 butir soal (40%) tidak memenuhi kriteria dengan rincian butir soal nomor 1, 3, 5, 8, 12, 18, 20, 21, 25, 26, 27, 28, 30, 35, 36, dan 39. Hasil ini menunjukkan bahwa kualitas tes UASBN IPA SD tahun pelajaran 2007/2008 menurut analisis teori tes klasik dikategorikan kurang baik.
3. Analisis kuantitatif (empiris) butir soal dengan pendekatan teori respons butir model Rasch menggunakan program *BIGSTEPS* diperoleh fungsi informasi tes 0,838, informasi maksimum 17,06 pada level kemampuan 0 (nol) logit dengan *SEM* 0,242 yang berarti tes ini handal. Ditemukan sebanyak 3 butir soal yang tidak cocok dengan model dan 2 butir soal yang memiliki kesukaran $b_i < -2$ atau sangat mudah. Secara keseluruhan terdapat 5 butir soal yang termasuk kriteria tidak baik menurut teori respons butir. Rincian soal yang tidak baik adalah soal nomor 1, 5, 25, 36, dan 39. Persentase butir tidak baik sebesar 5% menunjukkan kualitas tes berdasarkan teori respons butir dikategorikan baik.
4. Estimasi indeks konsistensi hasil analisis terhadap kedua metode yang digunakan yakni metode tes klasik dan metode respons butir memberikan nilai penafsiran

korelasi sedang. Besarnya indeks korelasi adalah 0,462 yang berarti hasil analisis menggunakan kedua metode tersebut cukup konsisten.

Menyimak temuan di atas, diperoleh informasi jumlah butir soal yang baik menurut ketiga metode analisis yang digunakan sebanyak 19 butir (47,5%) yakni butir nomor 2, 6, 7, 10, 13, 14, 15, 16, 17, 19, 22, 24, 29, 31, 32, 33, 37, 38, dan 40. Butir soal yang tidak baik sebanyak 21 butir (52,5%) yakni nomor 1, 3, 4, 5, 8, 11, 12, 18, 20, 21, 23, 25, 26, 27, 28, 30, 34, 35, 36, dan 39.

Saran

Kualitas tes UASBN IPA SD khususnya yang ditetapkan oleh penyelenggara tingkat provinsi Sulawesi Tenggara masih rendah (kurang baik). Kegiatan peningkatan kemampuan dan keterampilan guru SD dalam bentuk pendidikan dan latihan (Diklat), workshop, dan lokakarya sebaiknya diprogramkan oleh pihak Diknas provinsi maupun Diknas kabupaten/kota. Kerjasama pihak Dinas Pendidikan Nasional (Diknas), LPMP, dan perguruan Tinggi yang berkompeten dalam penyediaan bank soal sangat diperlukan. Apabila bank soal belum dapat diadakan oleh daerah, seyogyanya dalam proses penyusunan soal untuk keperluan UASBN dimasa yang akan datang dapat melibatkan ahli bidang studi, ahli bahasa, dan ahli pengukuran/psikometri sebagai tim pendamping/penelaah untuk menjaga mutu soal yang dihasilkan.

Daftar Pustaka

Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory*. Belmont, CA: Brooks/Cole Publishing Company.

- Bauer, S.C. (17 September 2000). Should achievement test be used to judge school quality. *Education policy analysis archives*, 46, 1-18.
- Djaali. (2006). *Hasil belajar evaluasi dalam evaluasi pendidikan: Konsep dan aplikasi*. Jakarta: Uhamka Press.
- Djemari Mardapi. (2008). *Teknik penyusunan instrument tes dan nontes*. Yogyakarta: Mitra Cendekia Press.
- Ebel, R. L. (1979). *Essential of educational measurement (3rd ed)*. New Jersey: Prentice-Hall, Inc., Englewood Cliffs.
- Gronlund, N.E. (1982). *Constructing achievement test. (3rd ed)*. New York: Prentice Hall, Inc., Englewood Cliffs.
- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R.K., & Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. London: Sage Publication.
- Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response theory. Application to psychological measurement*. Homewood: Dow Jones-Irwin.
- Ivailo Partchev. (2004). *A visual guide to item theory*. Friedrich-schiller: Universitat Jena.
- Kaplan, R.M & Saccuzzo, D.P. (1982). *Psychological testing. Principles, application, and issues*. California: Wadsworth, Inc.
- Kusmiyati. (2005). *Karakteristik butir tes ujian akhir IPA SLTP buatan guru*. Tesis magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- Naga, D.S. (1992). *Pengantar teori tes pada pengukuran pendidikan*. Jakarta: Gunadarama.
- Reid, C.A., et.al. (2007). Modern psychometric methodology: Applications of Item Response Theory [versi elektronik]. *ProQuest education journal*, 50, 177-189

- Saifuddin Azwar (2005). *Dasar-dasar psikometri (Edisi 1)*. Yogyakarta: Pustaka Pelajar.
- Santoso Murwani. (2006). *Evaluasi pendidikan: suatu pengantar, dalam evaluasi pendidikan: Konsep dan aplikasi*. Jakarta: Uhamka Press.
- Schumacker, R.E. (2005). Classical test analysis. Diambil pada tanggal 30 Agustus 2008, dari <http://www.appliedmeasure.mentassociates.com>.
- Sinharay, S., & Almond, R.G. (2007). Assessing fit of cognitive diagnostic models [versi elektronik]. *Journal educational and psychological measurement*, 67, (2), 239-257.
- Sumarna Surapranata.(2006). *Analisis, validitas, reliabilitas dan interpretasi hasil tes. Implementasi kurikulum 2004*. Bandung: PT. Remaja Rosdakarya Offset.
- Thorndike, R.M. (2005). *Measurement and evaluation in psychology and education (7th ed)*. New Jersey: Pearson Education, Inc.
- Tim Puslitbang Sisjian. (1999). *Pengantar model rasch dalam pengelolaan pengujian bagi guru mata pelajaran*. Jakarta: Ditjen Dikdasmen, Direktorat Pendidikan Menengah Umum.

Biodata Penulis

Muh. Nurung, M.Pd (Staf LPMP Sulawesi Tenggara)

Lahir di Liu, 25 Agustus 1972.

Pendidikan S1 Program Studi Pendidikan Fisika dari Fakultas Keguruan dan Ilmu

Pendidikan Universitas Haluoleo (UNHALU) Kendari tahun 2005; S2 Program Studi

PEP Pascasarjana Universitas Negeri Yogyakarta (UNY) tahun 2008.

