

PENGEMBANGAN BUTIR SOAL MATEMATIKA SD DI KABUPATEN LOMBOK TIMUR SEBAGAI UPAYA DALAM PENGADAAN BANK SOAL

Nila Hayati dan Djemari Mardapi
STKIP Hamzanwadi Selong NTB
email: hayatisyahdani@gmail.com

Abstrak

Penelitian ini bertujuan untuk menghasilkan butir-butir soal yang berkriteria baik dan melakukan penyeteraan butir soal perangkat tes mata pelajaran Matematika SD yang disiapkan untuk bank soal. Model yang digunakan dalam penelitian ini adalah model prosedural. Sampel siswa berjumlah 258 orang yang ditentukan dengan menggunakan teknik *stratified random sampling*. Perangkat tes yang diujikan meliputi dua paket soal (MAT_1 dan MAT_2) yang masing-masing terdiri dari 40 butir soal. Data dianalisis secara kualitatif dan kuantitatif. Analisis secara kualitatif didasarkan pada kesesuaian butir soal dengan lembar telaah berdasarkan kriteria materi, konstruksi, dan bahasa. Analisis kuantitatif menggunakan program *Iteman 3.00* untuk analisis dengan teori tes klasik dan program *Bilog MG* untuk analisis dengan teori respons butir model dua parameter. Hasil analisis kuantitatif menunjukkan dari 40 butir soal perangkat tes MAT_1 diperoleh 28 butir soal, sedangkan perangkat tes MAT_2 diperoleh 24 butir soal yang memenuhi kriteria untuk dimasukkan ke dalam bank soal. Ke-52 butir soal dari perangkat tes MAT_1 dan MAT_2 diarsip secara digital dengan aplikasi komputer.

Kata Kunci: Bank Soal, Matematika, Sekolah Dasar

DEVELOPING A MATHEMATICS ITEM TEST FOR ELEMENTARY SCHOOLS IN EAST LOMBOK REGENCY AS EFFORT IN SUPPLYING AN ITEM BANK

Abstract

This study was aimed at finding out the characteristics of developed mathematics test items for elementary schools and to equate test items in mathematics test sets for elementary schools developed for a test item bank. This study was a development research using the procedural model. The sample stratified random sampling, involving 258 students as respondents. The test sets were two test sets (MAT-1 and MAT-2), each consisting of 40 items. The data were quantitatively analyzed, using the *Iteman 3.00* and quantitatively analyzed using the *Bilog MG* program for the analysis using item response theory with the two-parameter model. The results of the quantitative analysis show that of the 40 test items in MAT-1 and MAT-2, 28 and 24 satisfy the criteria for the inclusion in the test item bank. The 52 test items in MAT-1 and MAT-2 are digitally archived using a computer application.

Keywords: item bank, test item development, elementary-school mathematics

PENDAHULUAN

Penilaian merupakan bagian penting dalam penyelenggaraan pendidikan. Penilaian dilakukan untuk mengukur dan menilai tingkat pencapaian kompetensi yang ada dalam kurikulum. Penilaian juga digunakan untuk mengetahui kekuatan dan kelemahan dalam proses pembelajaran, sehingga dapat dijadikan dasar untuk pengambilan keputusan,

dan perbaikan proses pembelajaran yang telah dilakukan. Selain itu juga hasil dari penilaian bisa memberikan motivasi kepada peserta didik. Hal ini sesuai dengan yang dikemukakan oleh Mardapi (2008:5):

“Salah satu upaya untuk meningkatkan kualitas pembelajaran adalah dengan meningkatkan kualitas sistem penilaian. Sistem penilaian yang

baik akan mendorong pendidik untuk menentukan strategi mengajar yang baik dan memotivasi peserta didik untuk belajar yang lebih baik”.

Oleh karena itu diperlukan sistem penilaian yang baik, terencana dan berkelanjutan untuk menunjang keberhasilan proses pembelajaran.

Salah satu langkah untuk menunjang keberhasilan dari sistem penilaian adalah dengan menyiapkan alat atau instrumen berupa tes dan non tes yang memenuhi persyaratan tes yang berkualitas. Tes merupakan suatu alat untuk menentukan satu contoh perilaku dari seseorang. Selanjutnya menurut Mardapi (2008: 67): “tes merupakan sejumlah pertanyaan yang memiliki jawaban yang benar dan salah”. Dengan demikian tes merupakan suatu alat yang terdiri dari sejumlah pertanyaan yang memiliki jawaban benar atau salah untuk mengukur perilaku dan karakteristik seseorang.

Suatu tes yang digunakan dalam proses penilaian tentu harus berkualitas baik dan tes tersebut benar-benar mampu mengukur kemampuan siswa yang sebenarnya. Ada dua hal yang harus diperhatikan dalam pembuatan perangkat tes, yaitu valid dan reliabel. Azwar (2010: 173) mengemukakan, “suatu tes dikatakan mempunyai validitas tinggi apabila tes tersebut menjalankan fungsi ukurannya, atau memberikan hasil ukur yang tepat dan akurat sesuai dengan maksud dikenakannya tes tersebut”. Selanjutnya Azwar (2010: 180) menyatakan “suatu tes dikatakan reliabel jika dalam beberapa kali pengukuran pada subjek yang sama atau berbeda diperoleh hasil yang relatif sama”. Akan tetapi banyak sekali ditemukan tes buatan guru belum memenuhi kriteria validitas dan reliabilitas. Hasil penelitian yang dilakukan oleh Purnomo (2007)

dan Suparji (2010), menemukan bahwa sebagian tes buatan guru belum memenuhi persyaratan validitas dan reliabilitas.

Selain itu juga keterlibatan guru masih bervariasi dalam melaksanakan penilaian khususnya dalam pembuatan soal. Hasil penelitian yang dilakukan oleh Mardapi, dkk (1999: 1) tentang survei kegiatan guru dalam melaksanakan penilaian di kelas menunjukkan, “pada tingkat Sekolah Dasar sebanyak 48% guru terlibat pembuatan kisi-kisi, 54% dalam penulisan soal, dan 53% dalam perakitan soal”. Selanjutnya Safari (Mujimin, 2008: 4) mengatakan:

“Kegiatan ujian di sekolah hanya menjadi kegiatan musiman tanpa adanya perencanaan sistematis. Hal yang mendukung dugaan ini di antaranya belum diwujudkan rencana di bidang penulisan soal di setiap sekolah, personil yang dilibatkan menangani pembuatan soal tes juga kurang menguasai kaidah teknik pada penyusunan butir soal, ujian hanya sekedar aturan atau prosedur yang harus dilaksanakan, jika ujian telah terlaksana maka selesailah sudah tanpa ada keinginan untuk menganalisis mutu bahan ujian, apalagi memanfaatkan hasil analisis untuk keperluan pengajaran”.

Berdasarkan hasil penelitian dan pendapat ahli di atas dapat dikatakan bahwa kegiatan penilaian kurang mendapat perhatian dari guru. Guru-guru lebih fokus terhadap penyampaian materi terlebih pada materi yang dianggap sulit seperti matematika. Selain itu juga kurangnya pemantauan dari kepala sekolah. Hasil penelitian yang dilakukan oleh Mardapi dkk tahun 1999 menyimpulkan bahwa semua kepala sekolah menyatakan mendorong guru dalam melaksanakan kegiatan penulisan soal, kisi-kisi soal, analisis butir soal, dan kegiatan

remediasi namun dalam pelaksanaannya belum dipantau dengan baik.

Pemerintah telah banyak berusaha untuk meningkatkan kualitas proses pembelajaran yang salah satunya adalah dengan melaksanakan penilaian secara nasional yang dikenal dengan Ujian Nasional (UN). Hal ini juga tertuang pada Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan bahwa penilaian pendidikan pada jenjang pendidikan dasar dan menengah terdiri atas: (a) penilaian hasil belajar oleh pendidik; (b) penilaian hasil belajar oleh satuan pendidikan; (c) penilaian hasil belajar oleh pemerintah. Penilaian hasil belajar oleh pemerintah bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu. Salah satu penilaian hasil belajar yang dilakukan oleh pemerintah untuk jenjang sekolah dasar adalah pelaksanaan Ujian Nasional (UN).

Pelaksanaan Ujian Nasional (UN) khususnya pada Sekolah Dasar (SD) dilaksanakan secara serentak di seluruh Indonesia. Berdasarkan peraturan yang ditetapkan pada Prosedur Operasi Standar (POS) UN SD, pelaksanaan UN SD diserahkan kepada pemerintah provinsi dan kabupaten/kota. Namun, dalam soal-soal UN SD, pemerintah pusat menitipkan 25% soal, sedangkan 75% soal lainnya disiapkan oleh pemerintah daerah. Dengan adanya kebijakan ini pembuatan soal diharapkan bisa disesuaikan dengan kondisi wilayah dan kondisi sekolah mengingat kondisi geografis Indonesia yang berbeda-beda, sehingga tes benar-benar mengukur kemampuan peserta yang sebenarnya.

Salah satu mata pelajaran yang dimasukkan dalam UN SD adalah matematika. Matematika merupakan salah satu mata pelajaran yang dianggap paling sulit oleh sebagian siswa, sehingga seringkali

prestasi belajar matematika siswa cenderung lebih rendah dari mata pelajaran yang lain. Hal ini terlihat dari hasil UN khususnya di Kabupaten Lombok Timur. Nilai hasil UN SD untuk mata pelajaran matematika rata-ratanya paling rendah jika dibandingkan dengan mata pelajaran yang lain yaitu sebesar 6,60. Untuk Bahasa Indonesia dan IPA masing-masing 6,99 dan 7,14 (Hasil laporan UN SD Kabupaten Lombok Timur tahun 2012).

Melihat data hasil UN SD di Kabupaten Lombok Timur khususnya mata pelajaran matematika, tidak serta merta langsung menempatkan siswa sebagai “kambing hitam” dengan sebutan kurang menguasai materi. Alangkah baiknya jika memeriksa kembali pelaksanaan pembelajaran yang telah dilaksanakan oleh guru. Apakah pembelajaran sudah direncanakan dengan baik? Apakah pelaksanaan pembelajaran sudah dilaksanakan sesuai dengan yang direncanakan? Apakah soal yang digunakan benar-benar sudah mengukur kemampuan siswa yang sebenarnya?

Hasil penelitian yang dilakukan oleh Mardapi, dkk (1999) menyatakan bahwa salah satu penyebab hasil UN rendah terutama ketika masih bernama Ebtanas adalah kesalahan dari alat ukur yang digunakan. Demikian pula penggunaan tes untuk kepentingan ujian di tingkat sekolah maupun *classroom asesment* oleh guru juga masih kurang memadai. Hal ini juga didukung oleh berbagai penelitian lain seperti penelitian yang dilakukan oleh Teddy (2009) tentang karakteristik soal matematika UASBN di Kota Pontianak, yang menyatakan bahwa perangkat soal UASBN secara keseluruhan memiliki kualitas kurang baik setelah dilakukan analisis dengan program ITEMAN meskipun berdasarkan hasil telaah dikategorikan baik. Penelitian lain juga yang dilakukan oleh Effendi (2009) tentang

karakteristik soal matematika ujicoba UASBN diperoleh bahwa karakteristik soal berdasarkan hasil analisis kualitatif memiliki kategori tidak baik, sedangkan hasil analisis kuantitatif dengan menggunakan teori tes klasik perangkat soal uji coba tersebut dikategorikan baik. Dengan demikian jika mengacu pada hasil penelitian-penelitian yang telah diungkapkan di atas dapat disimpulkan bahwa, salah satu penyebab rendahnya hasil UN SD untuk mata pelajaran matematika khususnya di Kabupaten Lombok Timur adalah alat ukur yang digunakan belum menjalankan fungsinya secara optimal.

Terkait dengan adanya kebijakan otonomi pendidikan dan peraturan 75% soal UN SD dikelola oleh pemerintah daerah, sehingga pengembangan bank soal di daerah sangat diperlukan. Hasil studi yang dilakukan oleh Mardapi dkk tahun 2001 merekomendasikan bahwa untuk memperkuat sistem pengujian dalam era otonomi daerah perlu pengadaan bank soal di setiap daerah yang dibina langsung oleh Pusat Penilaian Pendidikan. Di Kabupaten Lombok Timur sendiri keberadaan bank soal belum tersedia yang ada hanya sekumpulan dari soal-soal tanpa harus diketahui karakteristik dari soal tersebut (*Hasil wawancara dengan supervisor, tanggal 23 Oktober 2012*).

Keberadaan bank soal tentunya mempermudah guru dalam penyusunan instrumen setiap kali akan melaksanakan ujian, sehingga tidak harus tergesa-gesa dalam pembuatan soal. Kualitas soal dalam bank soal tentu berkualitas baik dan bisa dipertanggung jawabkan, karena sudah melewati prosedur yang telah ditentukan. Keberadaan bank soal di Kabupaten Lombok Timur sangat diperlukan, mengingat sedang berlangsungnya kebijakan Ujian Bersama baik untuk UTS maupun UAS.

Hasil wawancara (*tanggal 23 Oktober 2012*) dengan salah seorang supervisor

pendidikan di wilayah Kabupaten Lombok Timur dapat diketahui bahwa, pembuatan soal untuk Ujian Bersama ini dilakukan secara *roling* antarkecamatan. Selain itu soal-soal yang digunakan dalam ujian bersama ini diperoleh dari buku-buku yang menyediakan seribu satu macam soal tanpa perlu melakukan analisis dahulu apakah soal tersebut sudah layak untuk digunakan atau tidak. Tidak adanya bank soal tentunya menjadi penyebab masalah tersebut. Berdasarkan hal tersebut timbul ketertarikan peneliti untuk mengembangkan bank soal untuk mata pelajaran matematika SD yang nantinya bisa dimanfaatkan untuk Ujian Bersama atau untuk kegiatan pengujian lainnya.

Permasalahan yang diajukan dalam penelitian ini adalah sebagai berikut: (1) bagaimanakah karakteristik butir soal pada mata pelajaran matematika SD yang dikembangkan dan disiapkan untuk bank soal?, (2) Bagaimanakah penyetaraan butir soal perangkat tes mata pelajaran Matematika SD yang dikembangkan untuk bank soal? Penelitian ini bertujuan untuk: (1) menghasilkan produk berupa butir-butir soal yang berkriteria baik yang dikembangkan dan disiapkan untuk bank soal, (2) melakukan penyetaraan butir soal perangkat tes mata pelajaran Matematika SD yang dikembangkan untuk bank soal.

Hasil penelitian ini diharapkan bermanfaat sebagai sumber informasi bank soal mata pelajaran matematika tingkat SD di Kabupaten Lombok Timur sehingga dapat dimanfaatkan bagi sekolah dan guru. Selain itu dapat digunakan sebagai masukan atau dasar penelitian di bidang pengembangan Bank Soal khususnya pada mata pelajaran matematika tingkat SD dan umumnya untuk berbagai mata pelajaran dan tingkat sekolah.

METODE

Penelitian ini merupakan penelitian pengembangan dengan menggunakan

model prosedural. Model prosedural merupakan model yang bersifat deskriptif menunjukkan langkah-langkah yang harus diikuti untuk menghasilkan produk.

Pengembangan bank soal matematika SD di Kabupaten Lombok Timur diawali dengan kegiatan perencanaan untuk menentukan jenis dan bentuk tes yang akan dikembangkan. Selanjutnya dilakukan penulisan kisi-kisi soal. Penulisan kisi-kisi soal pada bank soal yang dikembangkan mengacu pada kisi-kisi soal Ujian Nasional matematika SD tahun 2011/ 2012. Selanjutnya dilakukan penulisan butir soal berbentuk pilihan ganda sejumlah 40 butir berdasarkan kisi-kisi yang telah dikembangkan. Butir soal yang telah disusun selanjutnya ditelaah dari segi materi, konstruksi, dan bahasa oleh para ahli (*Expert Judgement*). Hasil telaah butir soal kemudian dijadikan dasar revisi dan pembuatan perangkat uji coba prapenelitian. Selanjutnya dilakukan uji coba prapenelitian yang bertujuan melihat keterbacaan soal, karakteristik butir soal, dan penentuan butir *anchor*. Untuk melihat karakteristik butir soal yang meliputi tingkat kesukaran, daya beda, dan keberfungsian pengecoh di analisis secara teori tes klasik dengan menggunakan program *Iteman 3.00*. Dari hasil analisis dengan *Iteman 3.00* selanjutnya dilakukan revisi untuk butir soal yang mempunyai karakteristik kurang baik.

Berdasarkan hasil telaah dan analisis secara klasik disusun dua perangkat tes yang diharapkan paralel dengan disertai butir soal penggandeng (*anchor*) yang bermanfaat untuk analisis penyetaraan. Perangkat soal untuk pengambilan data penelitian diberi kode MAT-1 dan MAT-2 yang disusun dari kisi-kisi yang sama. Data hasil penelitian pengembangan ini berupa lembar jawaban respon siswa terhadap perangkat soal yang diberikan. Data berupa respon tersebut selanjutnya dikalibrasi dengan menggunakan

program *Bilog MG* dan dipilih butir soal yang memenuhi syarat untuk masuk dalam bank soal. Pengarsipan bank soal dalam penelitian ini dilakukan secara digital dengan menggunakan program aplikasi komputer. Pada penelitian ini juga terdapat proses penyetaraan, hal tersebut disebabkan terdapat 2 perangkat tes dari kisi-kisi yang sama yang digunakan dalam penelitian.

Subjek coba dalam penelitian ini adalah siswa kelas VI SD yang berjumlah 258 orang. Penentuan sampel dilakukan dengan teknik *stratified random sampling* berdasarkan hasil UN tahun 2011/2012 untuk mata pelajaran matematika. Hasil penyampelan ini diperoleh 2 sekolah dengan kategori rendah, 4 sekolah dengan kategori sedang, dan 2 sekolah dengan kategori tinggi.

Jenis data dalam penelitian ini berupa data hasil telaah ahli, data hasil prapenelitian, dan data hasil penelitian. Adapun instrumen yang digunakan dalam pengambilan data berupa lembar telaah naskah soal yang dikemas dalam dua paket yaitu MAT_1 dan MAT_2, serta lembar jawaban.

Teknik analisis data dalam penelitian ini dilakukan secara kualitatif dan kuantitatif. Analisis secara kualitatif terkait dengan penelaahan butir soal yang terkait dengan aspek materi, bahasa/budaya, konstruksi, dan kunci jawaban. Analisis secara kuantitatif menggunakan pendekatan teori tes klasik dengan bantuan program *Iteman 3.00* dilakukan untuk menganalisis hasil uji coba prapenelitian. Berdasarkan statistik butir soal akan diperoleh karakteristik tes meliputi tingkat kesukaran, daya beda, keberartian pengecoh. Selanjutnya dari hasil analisis butir akan diketahui karakteristik butir soal yang intinya apakah soal diterima, direvisi, atau ditolak. Tabel 1 merupakan kriteria pemilihan soal untuk pilihan ganda dari hasil analisis secara klasik.

02Selain menggunakan pendekatan teori tes klasik, analisis butir soal secara

Tabel 1. Kriteria Pemilihan Soal Pilihan Ganda Secara Teori Tes KI

Kriteria	Koefisien	Keterangan
Tingkat Kesukaran	0,30 s.d 0,70	Baik
Daya Beda	> 0,3	Baik
Proporsi Jawaban	>0,05	Baik

kuantitatif dalam penelitian ini juga menggunakan pendekatan teori tes modern atau teori respons butir. Pendekatan ini digunakan untuk menganalisis hasil uji coba penelitian dengan bantuan program *Bilog MG*. Seperti halnya dengan pendekatan teori tes klasik, dalam pemilihan soal dengan menggunakan teori respon butir juga memiliki kriteria seperti yang disajikan dalam Tabel 2.

Tabel 2. Kriteria Butir Soal Berdasarkan Teori Respons Butir

Model	Kriteria	
	Baik	Cukup Baik
IPL	$P > 0.05$ -2 b 2	$P > 0.05$ b < -2 atau b > 2
2PL	$P > 0.05$ 0 a 2; -2 b 2	Jika salah satu kriteria soal tidak terpenuhi
3PL	$P > 0.05$ 0 a 2; -2 b 2 0 c 1	Jika salah satu kriteria soal tidak terpenuhi

HASIL PENELITIAN DAN PEMBAHASAN

Telaah terhadap butir soal Matematika SD di Kabupaten Lombok Timur dilakukan oleh 2 (dua) orang penelaah yang memiliki kompetensi di bidang Matematika dan bidang pengukuran. Hasilnya menunjukkan bahwa dari 40 butir yang dikembangkan terdapat 21 soal yang diterima tanpa harus revisi dan 19 butir yang harus direvisi.

Hasil Analisis secara Tes Klasik

Analisis secara kuantitatif dengan pendekatan teori tes klasik dilakukan pada

data hasil prapenelitian dengan bantuan program *Iteman 3.00*. Berdasarkan hasil analisis diperoleh butir yang baik sebanyak 23 butir (55%) yakni nomor 2, 4, 6, 7, 9, 12, 13, 14, 16, 17, 19, 24, 25, 26, 28, 29, 30, 31, 33, 34, 35, 38, 39. Soal yang direvisi sebanyak 17 butir (45%) yaitu nomor 1, 3, 5, 8, 10, 11, 15, 18, 20, 21, 22, 23, 27, 32, 36, 37, 40. Indeks reliabilitas tes sebesar 0,834 dengan kesalahan baku pengukuran (SEM) sebesar 2,676.

Selain bertujuan untuk mengetahui karakteristik butir soal yang berupa tingkat kesukaran, daya beda dan distribusi jawaban, hasil analisis prapenelitian digunakan juga sebagai acuan untuk menentukan butir yang akan dijadikan sebagai butir *anchor* untuk keperluan proses penyetaraan. Jumlah butir yang dijadikan sebagai sebanyak 8 butir soal. Hal ini mengacu pada pendapat Skaggs & Lissitz (Sukirno 2007:34), yang menyatakan bahwa “jumlah butir jangkar (*anchor*) yang digunakan minimal 20% dari jumlah butir soal”. Untuk penulisan soal *anchor* mempunyai perbandingan komposisi 2:4:2, yaitu 2 butir soal dengan kategori mudah, 4 butir soal dengan kategori sedang, dan 2 butir soal dengan kategori sulit. Hal ini sesuai dengan yang dikemukakan Livingston (2004, 38-39) yaitu pertanyaan yang dimasukkan dalam butir *anchor* harus menggambarkan secara lengkap tingkat kesukaran dari butir soal, dikarenakan hasil *equiting* tidak bisa tepat jika hanya memasukkan soal yang memiliki tingkat kesulitan tinggi atau soal yang memiliki tingkat kesulitan rendah. Tabel 3 menyajikan sebaran butir yang akan dijadikan sebagai butir *anchor*.

Hasil Analisis secara Modern (Teori Respons Butir)

Analisis secara kuantitatif dengan pendekatan teori respon butir dengan bantuan program *Bilog MG* dilakukan

Tabel 3. Sebaran Butir *Anchor* Berdasarkan Hasil Analisis Uji Coba Pra-penelitian

Nomor Butir	Materi	Tingkat Kesukaran
Butir 1	Operasi hitung campuran bilangan cacah	Mudah
Butir 11	KPK dan FPB	Mudah
Butir 13	Perbandingan dan Skala	Sedang
Butir 18	Satuan berat	Sulit
Butir 32	Luas bangun gabungan	Sulit
Butir 35	Koordinat	Sedang
Butir 38	Penafsiran dan penyajian data	Sedang
Butir 39	Statistik diskriptif sederhana	Sedang

secara dua tahap. Tahap pertama dilakukan untuk keseluruhan data, sedangkan tahap kedua dilakukan untuk sebagian

data yang *valid* (korelasi biserial > 0,3) sesuai hasil analisis tahap pertama. Berdasarkan hasil analisis tahap pertama untuk perangkat tes MAT_1 diperoleh 29 butir, sedangkan perangkat tes MAT_2 diperoleh 26 butir. Hasil secara terperinci dapat dilihat pada Tabel 4 dan 5. Butir-butir tersebut selanjutnya dianalisis dengan menggunakan model 2 parameter yang meliputi 3 aspek yaitu tingkat kesukaran butir, daya beda, dan fungsi informasi.

Hasil analisis dengan program *Bilog MG* dapat diketahui bahwa pada perangkat tes MAT_1 terdapat 29 butir yang memiliki tingkat kesukaran dan daya beda berkategori baik dengan rerata berturut-turut 0,598 dan 0,473. Selain itu juga diperoleh

Tabel 4. Karakteristik Butir Soal dan Kecocokan Model Perangkat Tes MAT_1

No. Butir	Daya Beda		Tingkat Kesukaran		Fit Model	Keputusan
	Indeks	Ket	Indeks	Ket		
1	0,612	baik	0,038	baik	Fit	Diterima
2	0,487	baik	1,587	baik	Fit	Diterima
3	0,542	baik	0,358	baik	Fit	Diterima
4	0,616	baik	0,574	baik	Fit	Diterima
5	0,533	baik	0,731	baik	Fit	Diterima
6	0,481	baik	0,705	baik	Fit	Diterima
7	0,493	baik	1,804	baik	Fit	Diterima
8	0,866	baik	0,205	baik	Fit	Diterima
9	0,723	baik	0,713	baik	Fit	Diterima
10	0,663	baik	0,435	baik	Fit	Diterima
11	0,474	baik	-0,107	baik	Fit	Diterima
12	0,561	baik	1,131	baik	Fit	Diterima
13	0,701	baik	0,586	baik	Fit	Diterima
15	1,051	baik	-0,057	baik	Fit	Diterima
16	0,545	baik	1,67	baik	Fit	Diterima
17	0,452	baik	1,186	baik	Fit	Diterima
20	0,382	baik	1,738	baik	Fit	Diterima
23	0,848	baik	-0,412	baik	Tidak Fit	Direvisi
24	0,544	baik	-0,179	baik	Fit	Diterima
25	0,601	baik	0,472	baik	Fit	Diterima
27	0,794	baik	-0,894	baik	Fit	Diterima
28	0,647	baik	0,625	baik	Fit	Diterima
29	0,74	baik	1,121	baik	Fit	Diterima
31	0,556	baik	0,83	baik	Fit	Diterima
33	0,481	baik	1,226	baik	Fit	Diterima
34	0,846	baik	0,572	baik	Fit	Diterima
35	0,563	baik	-0,55	baik	Fit	Diterima
38	0,611	baik	0,692	baik	Fit	Diterima
39	0,656	baik	0,545	baik	Fit	Diterima

Tabel 5. Karakteristik Butir Soal dan Kecocokan Model Perangkat Tes MAT_2

No. Butir	Daya Beda		Tingkat Kesukaran		Fit Model	Keputusan
	Indeks	Ket	Indeks	Ket		
1	0,676	Baik	-0,209	baik	Fit	Diterima
2	0,977	Baik	-0,422	baik	Fit	Diterima
3	0,556	Baik	1,134	baik	Fit	Diterima
4	0,501	Baik	1,126	baik	Fit	Diterima
5	0,585	Baik	1,673	baik	Fit	Diterima
6	0,666	Baik	1,13	baik	Fit	Diterima
8	0,815	Baik	-0,013	baik	Fit	Diterima
9	0,4	Baik	1,058	baik	Fit	Diterima
10	0,903	Baik	0,71	baik	Fit	Diterima
11	0,778	Baik	0,152	baik	Fit	Diterima
12	0,542	Baik	0,472	baik	Fit	Diterima
13	0,664	Baik	0,399	baik	Fit	Diterima
14	0,721	Baik	-0,071	baik	Fit	Diterima
18	0,732	Baik	1,017	baik	Fit	Diterima
19	0,5	Baik	2,242	tidak baik	Fit	Direvisi
25	0,743	Baik	0,364	baik	Fit	Diterima
26	0,474	Baik	0,576	baik	Fit	Diterima
27	0,511	Baik	0,277	baik	Fit	Diterima
28	0,434	Baik	1,755	baik	Fit	Diterima
29	0,849	Baik	0,328	baik	Fit	Diterima
31	0,942	Baik	-0,427	baik	Fit	Diterima
33	0,546	Baik	0,22	baik	Fit	Diterima
35	0,576	Baik	-0,144	baik	Fit	Diterima
37	0,682	Baik	-0,934	baik	Tidak Fit	Direvisi
38	1,03	Baik	-0,641	baik	Fit	Diterima
39	0,623	Baik	0,538	baik	Fit	Diterima

informasi bahwa butir yang cocok (*fit*) dengan model sebanyak 28 butir yakni butir nomor 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 20, 24, 25, 27, 28, 29, 31, 33, 34, 35, 38, dan 39, sedangkan butir yang tidak cocok (*fit*) dengan model sebanyak 1 butir yakni butir nomor 23.

Hasil analisis dengan program *Bilog MG* pada perangkat tes MAT_2 dari 26 butir yang dianalisis diperoleh 25 butir memiliki tingkat kesukaran baik dengan rerata 0,623 dan 26 butir memiliki daya beda dengan kategori baik dengan rerata sebesar 0,670. Jumlah butir yang cocok (*fit*) dengan model

pada perangkat tes MAT_2 sebanyak 25 butir yakni butir nomor 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 18, 19, 25, 26, 27, 28, 29, 31, 33, 35, 37, 38, dan 39.

Fungsi Informasi dan SEM

Pada teori tes klasik koefisien reliabilitas digunakan untuk menilai mutu perangkat tes yang terdiri atas sejumlah butir, sedangkan pada teori respons butir untuk mencapai tujuan yang sama menggunakan fungsi informasi. Pada teori respons butir dikenal fungsi informasi butir dan fungsi informasi tes. Secara matematis fungsi

informasi parameter menurut Hambleton, Swaminathan & Rogers (1991: 91) dinyatakan dalam persamaan berikut.

$$I_i(\theta) = \frac{P_i(\theta)^2}{P_i(\theta)Q_i(\theta)} I_i(\theta) = \frac{P_i(\theta)^2}{P_i(\theta)Q_i(\theta)} \quad (1)$$

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir (Reid et al, 2007: 180). Hal ini berarti fungsi informasi tes akan tinggi jika butir soal penyusunnya mempunyai fungsi informasi yang tinggi pula. Fungsi informasi tes secara matematis dapat ditulis sebagai berikut.

$$I(\theta) = \sum_i^n I_i(\theta) \quad (2)$$

Pada teori respons butir kesalahan pengukuran disebut dengan *SE* (*Standar Error of Estimation*). Pada teori respons butir *SE* berkaitan erat dengan fungsi informasi. Fungsi informasi dengan *SE* mempunyai hubungan yang berbanding terbalik kuadratik. Makin besar nilai fungsi informasi berarti *SE* semakin kecil dan sebaliknya. Jika, fungsi informasi dinyatakan dengan $I(\theta)$ dan kesalahan baku pengukuran dinyatakan dengan $SE(\theta)$, bentuk hubungan keduanya dirumuskan sebagai berikut (Hambleton, Swaminathan, & Rogers, 1991: 94).

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (3)$$

Hasil perhitungan fungsi informasi dan SEM untuk perangkat tes MAT_1 dan perangkat tes MAT_2 disajikan pada Tabel 6 dan 7.

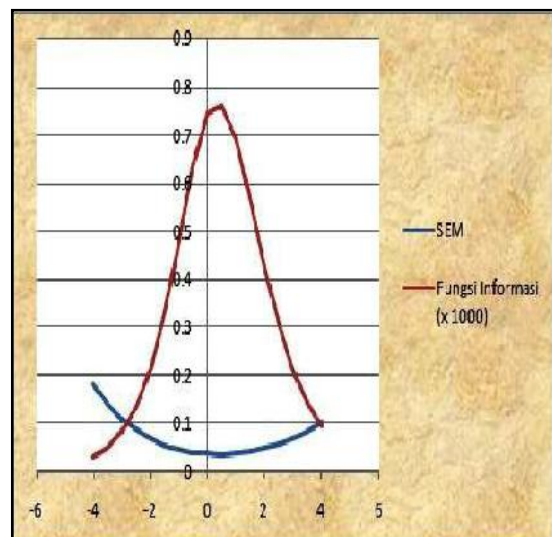
Lengkungan fungsi informasi dan SEM perangkat tes MAT_1 dapat dilihat pada Gambar 1.

Berdasarkan kurva Gambar 1 terlihat bahwa kurva membentuk lengkungan yang tajam. Hal ini berarti perangkat tes MAT_1 cenderung lebih akurat jika digunakan

Tabel 6. Harga Fungsi Informasi Tes dan SEM untuk Perangkat Tes MAT_1

Kemampuan		
()	I ()	SEM
-4	31,0278	0,180
-3,5	50,697	0,140
-3,0	82,893	0,110
-2,5	134,720	0,086
-2,0	215,076	0,068
-1,5	331,255	0,055
-1,0	480,625	0,046
-0,5	637,413	0,040
0,0	746,966	0,037
0,5	761,943	0,036
1,0	687,276	0,038
1,5	562,059	0,042
2,0	426,529	0,048
2,5	306,791	0,057
3,0	212,791	0,069
3,5	144,086	0,083
4,0	96,045	0,102

untuk mengukur kemampuan peserta pada tingkat kemampuan antara 0,0 sampai 0,5. Hal ini juga dibuktikan dengan nilai SEM yang paling kecil pada rentang kemampuan antara 0,0 sampai 0,5.

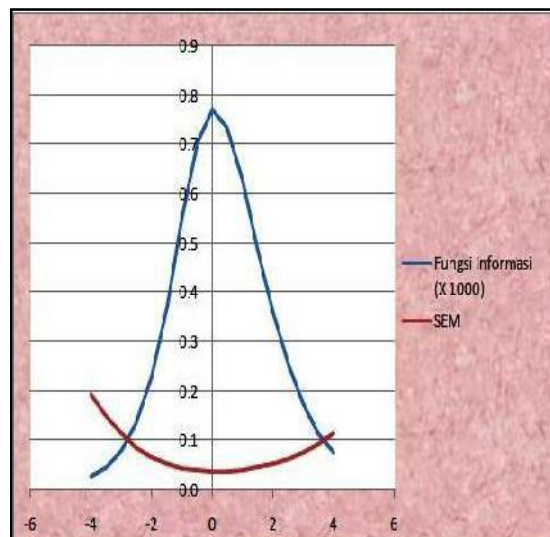


Gambar 1. Kurva Fungsi Informasi dan SEM Perangkat Tes MAT_1

Tabel 7. Harga Fungsi Informasi Tes dan SEM untuk Perangkat Tes MAT_2

Kemampuan		
()	I ()	SEM
-4,0	26,854	0,193
-3,5	45,727	0,148
-3,0	78,440	0,113
-2,5	134,694	0,086
-2,0	228,333	0,066
-1,5	371,261	0,052
-1,0	550,712	0,043
-0,5	705,759	0,038
0,0	769,561	0,036
0,5	733,266	0,037
1,0	626,681	0,04
1,5	489,765	0,045
2,0	358,647	0,053
2,5	251,900	0,063
3,0	172,363	0,076
3,5	115,917	0,093
4,0	77,021	0,114

Lengkungan fungsi informasi dan SEM perangkat tes MAT_2 dapat dilihat pada Gambar 2.



Gambar 2. Kurva Fungsi Informasi dan SEM Perangkat Tes MAT_2

Berdasarkan kurva Gambar 2 terlihat bahwa kurva membentuk lengkungan yang tajam. Hal ini berarti perangkat tes MAT_2 cenderung lebih akurat jika digunakan untuk mengukur kemampuan peserta pada tingkat kemampuan antara -0,5 sampai 0,5. Hal ini juga dibuktikan dengan nilai SEM yang paling kecil pada rentang kemampuan antara -0,5 sampai 0,5.

Penyetaraan (Equating)

Selanjutnya dilakukan proses penyetaraan untuk karakteristik butir dan kemampuan peserta. Proses penyetaraan menggunakan pendekatan teori respons butir dengan metode rerata sigma. Tabel 8 menyajikan tahapan penyetaraan penyetaraan tes dan kemampuan peserta dengan metode rerata sigma.

Tabel 8. Hasil Perhitungan Rerata dan Simpangan Baku Butir Anchor

Nomor	Nama	Tingkat Kesukaran		Daya Beda	
		MAT_1	MAT_2	MAT_1	MAT_2
1	Anchor 1	0,038	-0,209	0,612	0,676
11	Anchor 2	-0,107	0,152	0,474	0,778
13	Anchor 3	0,586	0,399	0,701	0,664
29	Anchor 4	1,121	0,328	0,740	0,849
35	Anchor 5	-0,55	0,144	0,563	0,576
38	Anchor 6	0,692	-0,641	0,611	1,03
39	Anchor 7	0,545	0,538	0,656	0,623
Mean		0,332	0,102	0,622	0,742
Standar Deviasi		0,566	0,405	0,088	0,156

Penyetaraan Butir Tes

Penyetaraan butir tes dengan menggunakan metode rerata sigma melalui beberapa tahapan sebagai berikut.

Menentukan *mean* dan *standar deviasi* tingkat kesukaran dan daya beda butir *anchor* dari masing-masing perangkat tes.

Penyetaraan Kemampuan Peserta

Selain penyetaraan skala pada parameter ciri butir, nilai konstanta penskalaan

dan dapat digunakan untuk menyamakan skala pada kemampuan peserta tes. Sebelum disamakan skalanya, parameter kemampuan peserta kelompok 1 yang mengerjakan tes MAT_1 adalah x dan parameter kemampuan peserta kelompok 2 yang mengerjakan tes MAT_2 adalah y .

Secara tes klasik kemampuan dapat dilihat dari skor peserta terhadap butir yang dijawab benar, sedangkan menurut teori respons butir kemampuan peserta dapat ditampilkan dalam bentuk skala dengan

satuan logit. Agar skala kemampuan ini dapat dimengerti perlu adanya transformasi ke bentuk lain yaitu umum dikenal dengan skala WITS.

Jika diperhatikan hasil konversi pada Tabel 9, diketahui rerata kemampuan peserta tes yang mengerjakan perangkat tes MAT_1 lebih besar dari peserta yang mengerjakan perangkat tes MAT_2. Hal ini berarti perangkat tes MAT_1 lebih sulit dari perangkat tes MAT_2.

Tabel 9. Hasil Konversi Kemampuan Peserta pada Skala yang Sama Sebelum dan Sesudah Ditransformasikan Ke Skala WITS

Skor	x	y^*	WITS	WITS*
0	-2.1805	-1.69506	80	85
2	-1.6805	-1.33756	85	88
3	-1.2136	-1.00372	89	91
4	-1.1078	-0.92808	90	92
5	-0.979	-0.83599	91	92
6	-0.8524	-0.74547	92	93
7	-0.7026	-0.63836	94	94
8	-0.3939	-0.41764	96	96
9	-0.158	-0.24897	99	98
10	0.0518	-0.09896	100	99
11	0.0275	-0.11634	100	99
12	-0.0236	-0.15287	100	99
13	0.2611	0.050687	102	100
14	0.4023	0.151645	104	101
15	0.5917	0.287066	105	103
16	0.6002	0.293143	105	103
17	0.7784	0.420556	107	104
18	0.8627	0.480831	108	104
19	1.0456	0.611604	110	106
20	1.3714	0.844551	112	108
21	1.3916	0.858994	113	108
22	1.6056	1.012004	115	109
23	1.7904	1.144136	116	110
24	1.9381	1.249742	118	111
25	1.9476	1.256534	118	111
26	2.2363	1.462955	120	113
29	2.9525	1.975038	127	118
Rerata			104	101

Pengaturan Butir Soal pada Bank Soal

Butir soal yang diterima dan siap dimasukkan ke dalam bank soal diarsip secara digital dengan menggunakan aplikasi komputer. Soal-soal yang ada pada kedua perangkat tes digabung menjadi satu kesatuan sehingga banyak jumlah soal ada 52 soal. Butir soal yang dimasukkan ke dalam bank soal diklasifikasikan berdasarkan kelas, sub mata pelajaran, topik bahasan, tingkat kesukaran, dan kemampuan peserta tes.

SIMPULAN

Berdasarkan analisis data dan pembahasan, simpul penelitian ini sebagai berikut. *Pertama*, hasil analisis perangkat tes MAT_1 yang terdiri dari 40 butir soal terdapat 28 butir soal yang berkarakteristik baik dan siap dimasukkan dalam bank soal dan 1 butir soal yang direvisi. Untuk perangkat tes MAT_2 dari 40 butir yang dianalisis terdapat 24 butir soal yang siap dimasukkan dalam bank soal dan 2 butir soal yang direvisi. Rerata indeks kesukaran perangkat tes MAT_1 adalah 0,598 dengan kategori baik dan rerata indeks kesukaran perangkat tes MAT_2 adalah 0,473 dengan kategori baik. *Kedua*, rerata indeks daya beda perangkat tes MAT_1 adalah 0,623 dengan kategori baik. Rerata indeks daya beda adalah 0,670 dengan kategori baik. *Ketiga*, nilai fungsi informasi tertinggi pada perangkat tes MAT_1 sebesar 761,943 dengan SEM 0,036 pada kemampuan () = 0,5, sedangkan pada perangkat tes MAT_2 nilai fungsi informasi tertinggi sebesar 769,561 dengan SEM 0,036 pada kemampuan () = 0,0. *Keempat*, pengaturan butir soal yang dikembangkan diklasifikasikan berdasarkan kelas, sub mata pelajaran, topik bahasan, tingkat kesukaran, dan tingkat kemampuan. *Kelima*, rerata kemampuan peserta yang mengerjakan perangkat tes MAT_1 dalam

skala WITS adalah 101, sedangkan rerata kemampuan peserta yang mengerjakan perangkat tes MAT_2 dalam skala WITS adalah 100. Berdasarkan hasil penyetaraan parameter butir dan peserta dapat diketahui bahwa perangkat tes MAT_1 lebih sulit dibandingkan dengan perangkat tes MAT_2.

DAFTAR PUSTAKA

- Azwar, S. 2010. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka pelajar.
- Depdiknas. 2005. Peraturan Pemerintah RI Nomor 19, Tahun 2005, tentang Standar Nasional Pendidikan.
- Hambleton, R. K, Swaminathan, H. & Rogers, H. J. 1991. *Fundamentals of Item Response Theory*. London: Sage Publication.
- Livingston, S.A. 2004. *Equating Test Score (without IRT)*. Educational Testing Service.
- Mardapi, D., Kartowagiran, B., Subali, B. 2001. "Sistem Ujian Akhir dalam Otonomi Daerah". *Laporan Penelitian*. Kerjasama Pusat Pengujian Balitbang Depdiknas dengan Fakultas Teknik Universitas Negeri Yogyakarta.
- Mardapi, D; Kartowagiran, B, Nurcholis, dan Fauzan. 1999. "Evaluasi Penyelenggaraan Ebtanas". *Laporan Penelitian*. Kerjasama Pusbangsisjian Balitbang Dikbud dengan Lembaga Penelitian IKIP Yogyakarta.
- Mardapi, D. 2008. *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendekia
- Mujimin. 2008. "Kompetensi Guru dalam Menyusun Butir Soal pada Mata Pelajaran Bahasa Jawa di Sekolah Dasar". *Tesis*. Tidak diterbitkan. Universitas Negeri Semarang.
- Purnomo, A. 2007. "Kemampuan Guru dalam Merancang Tes Berbentuk

- Pilihan Ganda pada Mata Pelajaran IPS untuk Ujian Akhir Sekolah (UAS)". *Lembaran Ilmu Kependidikan*, 36,1-6.
- Reid, C. A., Kolakowsky-Hayner, S. A., Lewis, Allen N., Armstrong, Amy J. 2007. *Modern Psychometric Methodology: Applications of Item Response Theory*. ProQuest Educa-
- Sukirno. 2007. "Penyetaraan tes UAN: Mengapa dan Bagaimana?". *Cakrawala Pendidikan*, 3, 305-321.
- Suparji. 2010. "Kualitas Butir Soal Buatan Guru-Guru SMP Mata Pelajaran Matematika dan IPA Di Kabupaten Sumenep". *Jurnal Pendidikan Dasar*, 1, 48-52.