

PENYETARAAN SKOR TES FISIKA SMA DENGAN MODEL RASCH

Oleh:
Sigit Purwanto

Abstrak

Penelitian ini bertujuan: (1) membuat tiga perangkat tes fisika yang setara untuk SMA kelas 3 pada ujian akhir nasional (UAN), (2) menghitung persamaan konversi tingkat kesukaran dan kemampuan *testee* dari ketiga perangkat tes fisika.

Subjek uji adalah tiga kelompok siswa kelas 3 SMA Negeri di Kabupaten Bantul, yang berturut-turut berjumlah 99 siswa, 105 siswa, dan 102 siswa, sehingga seluruhnya berjumlah 306 siswa. Penentuan subjek penelitian dilakukan dengan teknik *proportional random sampling*. Data penelitian diperoleh melalui tes dan diolah menggunakan program Rascal untuk analisis hasil ujicoba tes dan program Bigstep untuk penyetaraan skor ketiga tes. Hasil telaah butir soal ketiga tes menunjukkan bahwa dari 60 butir soal jumlah butir soal yang baik adalah 52 untuk tes I, 51 untuk tes J, dan 55 untuk tes K, sedangkan selebihnya memerlukan revisi.

Hasil analisis ujicoba menunjukkan hal-hal sebagai berikut. Pertama, tingkat kesukaran sebagian besar butir soal dalam ketiga tes termasuk dalam kategori menengah. Kedua, jumlah butir soal yang valid dalam tes I adalah 47 butir, tes J adalah 44 butir, tes K adalah 46 butir, sedangkan selebihnya tidak valid. Urut-urutan tingkat kesukaran ketiga tes, dari yang paling sukar adalah: tes K, J, dan I, sehingga kemampuan *testee* yang paling tinggi dituntut oleh Tes K, diikuti tes J dan kemudian tes I. Hasil perhitungan persamaan konversi skor untuk penyetaraan antar tes yaitu: (1) Persamaan konversi tingkat kesukaran butir soal $bKI^* = bI - 0,408$ dan $bKI^* = bJ - 0,242$, (2) Persamaan konversi kemampuan *testee*: $0KI^* = 0I - 0,408$ dan $0KJ^* = 0J - 0,242$.

Kata kunci: *penyetaraan, fisika sma, model rasch.*

Pendahuluan

Proses Belajar Mengajar (PBM) di sekolah bertujuan agar siswa dapat menguasai pengetahuan yang dipelajari, bersikap positif serta trampil melakukan sesuatu yang dipelajarinya. Dengan demikian, tinggi rendahnya ketiga hal tersebut merupakan indikator yang menunjukkan tingkat keberhasilan proses belajar mengajar.

Untuk mengetahui tingkat penguasaan siswa atas materi pelajaran dan mengetahui keberhasilan siswa perlu dilakukan pengukuran. Kecermatan dan ketepatan cara pengukuran dan alat pengukuran yang digunakan dalam penilaian sangat menentukan keputusan dan kebijakan yang diambil dalam bidang pendidikan. Hasil pengukuran dan penilaian merupakan informasi yang tepat untuk melakukan tindakan lebih lanjut.

Untuk mengukur dan menilai prestasi belajar siswa dapat digunakan berbagai macam alat ukur, yang salah satu diantaranya adalah Ujian Akhir Nasional (UAN). Hasil ujian akhir ini sangat penting dalam pengambilan keputusan, khususnya dalam penentuan kelulusan siswa. Agar hasil ujian dapat menggambarkan keadaan sesungguhnya, maka pengukuran harus dilakukan secara profesional, akurat, objektif dan adil.

Ujian Akhir Nasional adalah kegiatan penilaian hasil belajar siswa yang telah menyelesaikan jenjang pendidikan yang diselenggarakan secara nasional yang sebelumnya disebut Evaluasi Belajar Tahap Akhir Nasional (EBTANAS). Pelaksanaan EBTANAS telah dilakukan beberapa tahun yang lalu, namun hasil yang diperoleh masih rendah. Hal ini ditunjukkan dengan perolehan Nilai Ebtanas Murni (NEM) yang rendah, khususnya mata pelajaran fisika. Berdasarkan data dari Kantor Dinas Pendidikan dan Kebudayaan Kabupaten Bantul, rata-rata NEM mata pelajaran fisika pada tahun pelajaran 2001/2002 yaitu 3,14. Mulai tahun pelajaran 2002/2003 mata pelajaran fisika diujikan di sekolah masing-masing, soal dibuat oleh guru masing-masing sekolah, sehingga soalnya tidak sama. Di samping itu tes yang digunakan tidak diketahui kualitasnya. Karakteristik butir tes tidak diketahui, sangat sukar atau sangat mudah. Kualitas soal hanya didasarkan pada kemampuan dan penilaian penyusun soal.

Dengan menggunakan tes yang tidak sama dan tidak diketahui karakteristiknya, maka hasil yang diperoleh siswa antarsekolah berbeda, sehingga tidak dapat dibandingkan atau digabungkan. Di samping itu tidak dapat diketahui perkembangan hasil yang diperoleh siswa pada masing-masing sekolah. Untuk mengetahui perkembangan mutu masing-masing sekolah maka perlu penyetaraan skor.

Penyetaraan perangkat tes ujian merupakan hal yang penting dalam kegiatan penilaian yang menggunakan lebih dari satu perangkat tes. Hal ini untuk memperoleh hasil yang *comparable* antara nilai hasil tes paket satu dengan paket yang lain, apalagi jika hasilnya digunakan sebagai salah satu dasar penentuan kelulusan.

Menurut Sumadi Suryabrata (1987: 12) dalam pelaksanaan penilaian yang menggunakan beberapa tes yang berbeda, perlu dilakukan penyetaraan dari tes tersebut, karena dengan penyetaraan dapat dijamin keadilan bagi peserta tes yang menempuh tes yang berbeda. Dengan demikian, dalam mengembangkan beberapa tes yang akan digunakan sebagai alat tes, terlebih dahulu dilakukan penyetaraan. Keuntungan dilakukan penyetaraan adalah:

1. Dimungkinkan penggunaan tes yang berbeda terhadap kelompok yang berbeda sesuai dengan tingkat kemampuannya, sehingga skor yang diperoleh dapat dibandingkan dan peserta tes (*testee*) tidak merasa dirugikan atau diuntungkan karena mereka kebetulan mengerjakan tes yang lebih sukar atau lebih mudah
2. Hasil ujian dapat dibandingkan dari tahun ke tahun serta dapat mengetahui perkembangan atau kemajuan prestasi yang dicapai
3. Bila terjadi kebocoran dari satu tes tertentu dapat segera diganti dengan tes lain yang telah disiapkan.

Dua teori tes yang berkaitan dengan penilaian hasil belajar, yaitu teori tes klasik dan teori respons butir (*item response theory*). Pengembangan teori respons butir dimaksudkan untuk mencoba mengatasi kelemahan-kelemahan yang dimiliki teori tes klasik. Teori tes klasik mempunyai beberapa kelemahan antara lain terletak pada keterkaitan antara tes dengan peserta tes, maksudnya kemampuan peserta tes ditentukan oleh tingkat kesukaran tes tersebut dan

sebaliknya kesukaran tes ditentukan oleh kemampuan peserta tes. Hal ini dapat diatasi dengan teori respons butir.

Penggunaan teori respons butir dalam pengujian, seperti ujian akhir sekolah pada mata pelajaran fisika, khususnya model Rasch (model satu parameter) memiliki kesederhanaan dan tidak mengurangi kualitas hasil pengukuran yang diperolehnya. Dengan memanfaatkan parameter tingkat kesukaran, pengelola pengujian dapat mengkonfirmasi butir-butir tes terkalibrasi sesuai dengan tujuan tes. Butir-butir soal yang invarian terhadap parameter butir dan peserta tes memberikan hasil estimasi kemampuan dan estimasi butir yang *reliable* dan tidak akan mengubah status butir tes, sekalipun butir tes tersebut dijawab oleh siapapun atau peserta tes menjawab butir soal mana saja.

Penelitian ini bertujuan untuk mengembangkan tiga soal ujian akhir fisika yang setara untuk tingkat SMA negeri kelas 3 pada semester 2, berdasarkan kisi-kisi yang sama. Kemudian diperolehnya karakteristik soal ujian akhir fisika yang sudah disetarakan, maka akan memberikan kesempatan yang merata dan adil bagi semua peserta, selain itu dengan dilakukan penyetaraan diperoleh persamaan konversi skor antartest fisika SMA negeri, sehingga skor yang diperoleh antartest dapat dibandingkan.

Penafsiran hasil pengukuran sering disebut dengan istilah penilaian, yaitu untuk menyatakan kualitas prestasi belajar siswa dan perkembangan mutu pendidikan. Penafsiran hasil pengukuran melibatkan penentuan subjektif kolektif, yaitu merupakan kesepakatan intersubjek. Selanjutnya apabila penafsiran ini diikuti dengan keputusan kebijakan, kegiatan ini sudah sampai taraf evaluasi. Oleh karena itu pengukuran, penilaian, dan evaluasi merupakan kegiatan yang berkaitan dan berurutan (Djemari Mardapi, 1999: 7).

Penilaian merupakan proses penggambaran, memperoleh dan memberikan informasi yang berguna sebagai alternatif pengambilan keputusan. Suatu proses penilaian ada yang didasarkan atas hasil pengukuran yang bersifat kuantitatif dan ada pula yang bersifat kualitatif yang didasarkan atas non pengukuran. Ciri pokok pengukuran adalah adanya proses perbandingan. Pengukuran memerlukan alat ukur, dan salah satu alat ukur yang digunakan adalah tes prestasi belajar.

Dalam hubungannya dengan pendidikan, Gronlund (1976: 6) berpendapat, bahwa penilaian dapat diartikan sebagai suatu proses yang sistematis untuk menentukan sejauh mana tingkat pencapaian para siswa terhadap tujuan-tujuan pendidikan yang telah ditetapkan. Batasan tersebut mengandung dua hal penting yakni, bahwa penilaian merupakan suatu proses yang sistematis, penilaian terdiri dari serangkaian kegiatan yang dilakukan yang dilalui dan berdasarkan aturan –aturan tertentu.

Menurut Hariwibowo (1994: 3) dan Depdikbud (1994: 4) penilaian terbagi atas:

1. Penilaian tingkat mikro, yaitu penilaian yang dilaksanakan pada saat interaksi dengan siswa di kelas.
2. Penilaian tingkat makro, yaitu penilaian berkenaan dengan mekanisme pengalaman program, seperti sumatif atau ujian akhir nasional.

Sehubungan dengan itu, hal penting yang harus diperhatikan dalam penilaian adalah sifatnya harus objektif dan adil. Objektif maksudnya sesuai dengan keadaan sesungguhnya atau apa adanya tanpa rekayasa. Sedangkan adil berarti bahwa penilaian dilakukan dengan memberi perlakuan yang sama kepada semua orang yang dinilai. Dengan demikian alat ukur yang digunakan hendaknya ditujukan untuk mengukur hal yang sama. Begitu juga mengenai data yang diperoleh diolah, diinterpretasikan dengan cara sama dan skala penilaian/pengukuran yang digunakan juga harus sama. Selain itu termasuk pengertian adil adalah memperhatikan adanya perbedaan kemampuan antara siswa sesuai dengan kondisi masing-masing sekolah. Untuk itu diperlukan informasi lebih lanjut mengenai kemampuan siswa.

Sumadi Suryabrata (1990: 327) mengemukakan, bahwa ujian itu adalah tes, sedangkan tes adalah suatu pengukuran objektif dan standar terhadap sampel perlakuan. Menurut Depdiknas (2000: 11) tes ialah himpunan pertanyaan yang harus dijawab, atau pertanyaan-pertanyaan yang harus dipilih atau ditanggapi atau tugas-tugas yang harus dilakukan oleh orang yang dites (*testee*) dengan tujuan untuk mengukur suatu aspek (perilaku) tertentu orang yang dites.

Teori Respons Butir

Menurut Hambleton & Swaminathan (1985: 13) teori respons butir dikembangkan atas dasar pemikiran: (a) hasil tes seseorang dapat diprediksi dari kemampuan yang dimilikinya, (b) hubungan antara hasil tes dan kemampuan dinyatakan dalam sebuah fungsi yang disebut kurve karakteristik butir atau *item characteristic curve/ICC*. Kurve karakteristik butir akan menunjukkan peluang seseorang yang memiliki kemampuan tertentu menjawab benar suatu butir soal pada tes. Model-model kurve karakteristik butir digambarkan dengan satu, dua, atau tiga parameter, bergantung pada model mana yang digunakan.

Teori tes modern atau *Item Respons Teori (IRT)* dikembangkan karena adanya keterbatasan pada teori pengukuran klasik yaitu bersifat *group dependent* dan *item dependent* (Hambleton, Swaminathan & Rogers, 1991: 3). *Group dependent* artinya pengukuran tergantung dari kelompok peserta yang mengerjakan, jika peserta tes dengan kemampuan tinggi akan memiliki peluang yang lebih besar untuk menjawab dengan benar terhadap suatu butir soal yang memiliki kesukaran tinggi, sebaliknya peserta tes dengan kemampuan rendah memiliki peluang yang lebih kecil untuk dapat menjawab dengan benar terhadap suatu butir soal yang memiliki tingkat kesukaran tinggi. Kemampuan di sini, yang juga sering disebut dengan potensi, merupakan faktor dominan untuk menentukan keberhasilan seseorang dalam belajar yang ditunjukkan oleh hasil yang diperolehnya dari tugas suatu tes. *Item dependent* artinya pengukuran tergantung dari tes mana yang diujikan, jika tes diujikan mempunyai kesulitan tinggi maka estimasi kemampuan peserta tes rendah, sedangkan jika tes yang diujikan mempunyai tingkat kesulitan rendah maka estimasi kemampuan tes tinggi.

Tujuan teori respons butir adalah mengembangkan suatu skala pengukuran yang seragam dan bersifat *sample free*. artinya alat ukur tersebut dapat dipakai untuk mengukur hasil belajar individu maupun kelompok siswa dengan kemampuan bervariasi. Dalam konsep teori ini, kemampuan merupakan bentuk statistik yang diturunkan secara matematis dari serangkaian jawaban soal, yang berbentuk dikotomi, yang diujikan kepada sekelompok siswa yang

jumlahnya cukup besar. Menurut Hambleton, Swaminathan & Rogers (1991: 5) tujuan teori respons butir adalah untuk: (a) ciri butir yang tidak tergantung pada kelompok subjek, (b) skor tes yang dapat menggambarkan kemampuan subjek dan tidak tergantung kelompok subjek yang mengerjakan, (c) model tes dapat memberikan pencocokan antara item tes dan tingkat kemampuan, (d) model tes yang tidak memerlukan asumsi paralel dalam pengujian reliabilitasnya, dan (e) model tes yang asumsi-asumsinya mempunyai dukungan yang kuat.

Menurut Dali S. Naga (1992: 176-192) terdapat beberapa model dalam teori respons butir, antara lain: model skala sempurna (model jarak laten), model linear, model ojaif menengah dan model logistik. Di antara model-model tersebut yang banyak digunakan adalah model logistik, karena pendekatannya lebih sederhana, dan lebih mudah dalam aplikasi analisis. Model logistik ini dapat dibedakan atas tiga macam, yaitu model tiga parameter, model dua parameter, dan model satu parameter yang dikenal dengan model Rasch (Djemari Mardapi, 1991: 8).

Penelitian ini menggunakan model logistik satu parameter atau model Rasch. Konsep yang dikembangkan oleh Rasch dalam mengembangkan model ini berbeda dengan konsep teori respons butir. Parameter butir pada model ini adalah tingkat kesukaran butir, sedangkan parameter daya pembeda dianggap sama dan peluang tebakan butir sama dengan nol.

Tingkat kesukaran butir soal merupakan fungsi dari kemampuan seseorang. Besarnya kemampuan seseorang diestimasi berdasarkan tanggapannya pada butir tes. Seseorang memiliki kemampuan tinggi akan merasa mudah untuk mengerjakan butir soal. Sebaliknya seseorang yang memiliki kemampuan rendah akan merasa sulit untuk mengerjakan butir soal. Pada dasarnya besarnya parameter butir soal adalah tetap, yaitu setelah melalui kalibrasi.

Penyetaraan Skor Tes

Menurut Kolen dan Brennan (1995: 28) sifat spesifikasi-spesifikasi yang sama adalah sebuah sifat utama bagi penyetaraan, yang berarti bahwa bentuk-

bentuk yang akan disetarakan harus dibangun menjadi isi spesifikasi-spesifikasi statistik yang sama.

Sumadi Suryabrata (1987: 45) menyatakan bahwa penyetaraan ialah usaha mengkonversikan sistem unit-unit suatu perangkat tes ke sistem unit-unit perangkat tes yang lain, sehingga skor-skor yang dijabarkan dari dua perangkat tes, setelah dikonversikan secara langsung akan menjadi setara. Allen dan Yen (1979: 170) mengemukakan bahwa pada saat membuat tes alternatif, jika mungkin dirasakan perlu untuk mentransformasikan skor tes, sehingga menjadi setara, inilah yang disebut penyetaraan skor. Petersen, Kolen & Hoover (Linn, 1989: 242) menyatakan, bahwa penyetaraan adalah suatu prosedur empiris dalam rangka menyetarakan skor tes dari satu tes ke tes yang lain, agar diperoleh skor yang setara.

Pendapat-pendapat di atas memberikan pengertian bahwa penyetaraan ialah suatu prosedur yang dilakukan dalam rangka membuat setara skor dari satu perangkat tes ke perangkat tes yang lain. Adapun tujuan dilakukannya penyetaraan ialah untuk mewujudkan sedekat mungkin skor dari dua perangkat tes.

Penyetaraan skor dapat dilakukan terhadap skor kemampuan (*ability score equating*), skor amatan (*observed score equating*) dan parameter butir. Dalam penelitian ini penyetaraan dilakukan terhadap parameter butir (tingkat kesukaran butir) dan skor kemampuan.

Penyetaraan yang dilakukan dalam penelitian ini menggunakan pendekatan teori respons butir dengan metode rerata dan sigma. Hambleton, Swaminathan & Rogers (1991: 129) berpendapat bahwa penyetaraan skor yang dilakukan dengan prosedur metode rerata dan sigma, hasil konversi skor tes I ke tes J adalah I^* , dan sebaliknya hasil konversi skor tes J ke tes I adalah J^* dengan formulasi sebagai berikut :

$$b_{vj} = \alpha b_{vi} + \beta \quad (3)$$

Keterangan:

b_{vj} dan b_{vi} = skor atau kesukaran butir yang disetarakan (tes I ke tes J)
 α dan β = konstanta yang dicari dari data tes *anchor*.

Estimasi parameter butir untuk butir inti dalam tes J tidak akan identik dengan estimasi parameter butir pada tes I, perbedaan itu disebabkan oleh kesalahan estimasi, oleh sebab itu untuk mengontrol perbedaan skor atau parameter butir tersebut dirata-rata. Dengan kata lain, skor atau tingkat kesukaran butir dari kedua tes *anchor* merupakan nilai konstanta b penyetaraan.

Skor *anchor* V yang diperoleh dari tes I menghasilkan skor V_I dan skor *anchor* yang diperoleh dari tes J menghasilkan skor V_J . Estimasi parameter butir rerata $\mu_{b_{VI}}$ dan standar deviasi δb_{VI} untuk tes *anchor* pada tes I dan rerata $\mu_{b_{VJ}}$ dan standar deviasi δb_{VJ} untuk *anchor* pada tes J, rerata $\mu_{b_{VK}}$ dan standar deviasi δb_{VK} untuk *anchor* pada tes K.

Konstan penskalaan adalah konstan yang terdapat dalam rumus penyetaraan. Penetapan nilai konstan dihitung dengan menggunakan bentuk matematika sebagai berikut :

$$b_{VJ} = \alpha b_{VI} + \beta \quad (4)$$

Metode rerata dan sigma ini memasukkan rerata dan standar deviasi ke dalam rumus penyetaraan. Pada standar deviasi, konstan translasi saling meniadakan, sehingga ditemukan bentuk :

$$\mu_{b_{VI}} = \alpha \mu_{b_{VJ}} + \beta \quad (5)$$

$$\text{dan } \alpha \delta b_{VI} = \alpha \delta b_{VJ} \quad (6)$$

Hubungan tersebut untuk menghasilkan rumus untuk mencari a dan b .

$$\alpha = \frac{\alpha b_{VI}}{\delta b_{VJ}} \quad (7)$$

$$\text{dan } \beta = \mu_{b_{VI}} - \alpha \mu_{b_{VJ}} \quad (8)$$

Dengan demikian diperoleh konstan penyetaraan a dan b sebagai konstan translasi. Hasil konstan a dan b dapat disubstitusikan ke dalam persamaan (7), sehingga akan menghasilkan persamaan konversi penyetaraan tes J ke tes I. Transformasi tingkat kesukaran dari tes I ke tes J dapat dilakukan dengan persamaan :

$$b_I = \frac{b_J - \beta}{\alpha} \quad (9)$$

Estimasi parameter kemampuan dan butir dapat ditentukan dalam skala yang sama dengan menggunakan hubungan :

$$\theta_{JI}^* = \alpha \theta_I + \beta \quad (10)$$

$$b_{JI}^* = \alpha b_I + \beta \quad (11)$$

$$a_J^* = \frac{a_I}{\alpha} \quad (12)$$

Keterangan :

θ_{JI}^* = kemampuan testee pada tes J sebagai kriterium,

θ_I = kemampuan testee pada tes I sebagai prediktor (yang disetarakan dengan tes J),

b_{JI}^* = skor atau tingkat kesukaran butir tes pada tes J sebagai kriterium.

α dan β = konstan yang dihitung dari data tes anchor, b_I = skor atau tingkat kesukaran tes I sebagai prediktor (yang disetarakan ke tes J),

a_J^* = diskriminasi butir tes J sebagai kriterium,

a_I = diskriminasi butir tes I sebagai prediktor.

Lambang θ , b dan a merupakan parameter pada ICC. Kalau parameter tersebut dimasukkan ke dalam skala lain yang disamakan, parameter tersebut menjadi θ^* , b^* dan a^* . Simbol α dan β merupakan konstanta.

Pada model logistik satu parameter $a^* = 1$, sehingga $\alpha = 1$. Hal ini menyebabkan rumus penyeteraan menjadi :

$$\theta_{JI}^* = \theta_I + \beta \quad (13)$$

$$b_{JI}^* = b_I + \beta \quad (14)$$

$$a_J = a_I \quad (15)$$

Metode rerata sigma mempunyai kelebihan, karena rumus penyetaraan dapat dilakukan secara bolak-balik, yaitu dari b_j ke b_i . Transformasi dari b_i ke b_j dapat diperoleh rumus :

$$b_i = \frac{b_j - \beta}{\alpha} \quad (16)$$

Metode rerata dan sigma memenuhi syarat simetri, rerata dan standar deviasi digunakan untuk menentukan a dan b . Selain penyetaraan skala parameter butir, nilai konstanta a dan b dapat juga digunakan untuk menyetarakan skala parameter kemampuan.

Langkah –langkah Penelitian:

1. Menyusun kisi – kisi tes fisika
2. Membagi kisi-kisi soal kepada 3 orang guru yang membuat soal (jumlah masing-masing 45 butir soal)
3. Soal yang sudah dibuat oleh guru sebanyak tiga perangkat soal dengan kisi-kisi yang sama, kemudian ditambahkan 15 soal *anchor* pada masing-masing perangkat soal tersebut
4. Terdapat 3 perangkat soal yaitu: perangkat tes I, perangkat tes J dan perangkat tes K
5. Ketiga perangkat soal ditelaah oleh 3 orang penelaah (penelaah bukan penyusun soal) yang berbeda.
6. Butir soal yang salah pada masing-masing perangkat, kemudian direvisi
7. Melakukan ujicoba ketiga perangkat soal.
8. Menganalisis hasil ujicoba dengan program Rascal.
9. Dari hasil analisis ujicoba diperoleh tiga perangkat soal dengan jumlah 40 butir soal.
10. Hasil ketiga perangkat soal yang diujikan dianalisis dengan program Bigsteps untuk kepentingan penyetaraan.
11. Rerata tes *anchor* yang diperoleh, digunakan untuk menentukan konstanta translasi.

Metode Penelitian

Variabel penelitian ini adalah tingkat kesukaran butir dan tingkat kemampuan *testee*. Tingkat kesukaran butir dari masing-masing perangkat tes kemudian digunakan untuk menyatakan urutan tingkat kesukaran dari ketiga tes. Tingkat kemampuan *testee* dari peserta yang mengerjakan masing-masing perangkat tes untuk menyatakan tingkat kemampuan dari penempuh ketiga perangkat tes.

Subjek uji dalam penelitian ini adalah siswa kelas 3 SMA Negeri Kabupaten Bantul Jurusan IPA yang sedang menempuh Ujian Akhir Nasional (UAN). Pengambilan subjek penelitian menurut wilayah dan jumlah subjek. Perincian subjek penelitian dari setiap SMA yang terambil tertera pada tabel berikut:

Tabel 1
Perincian subjek penelitian SMA Negeri wilayah POKJA 2
Kabupaten Bantul

Kelompok	Nama Sekolah	Banyak data	Subjek
A	SMA N 1 Jetis	99	99
B	SMA N SewonSMAN 1 Dlingo	2580	105
C	SMA N 1 Banguntapan	102	102

Untuk kepentingan penyeteraan, penelitian ini menggunakan tiga tes sebagai instrumen penelitian, yaitu tes I yang terdiri dari tes I dan V, tes J yang terdiri dari tes J dan tes V, dan tes K yang terdiri dari tes K dan tes V. Tes V adalah tes *anchor* yang ditempatkan berdasar tes I, tes J dan tes K. Ketiga tes tersebut dikembangkan pada kisi-kisi yang sama. Kisi-kisi tes disusun oleh peneliti sendiri, sesuai dengan Standar Kompetensi Lulusan dan Spesifikasi Ujian Akhir SMA/MA yang disusun oleh Pusat Penilaian Pendidikan, Badan Penelitian dan Pengembangan, Departemen Pendidikan Nasional Tahun 2004. Kemudian penyusunan tes dilakukan oleh tiga kelompok guru yang mengajar mata pelajaran fisika pada ketiga kelompok sekolah tersebut, satu kelompok

guru menyusun satu perangkat tes, sehingga terdapat tiga perangkat tes yang berbeda.

Sebelum dilakukan uji coba, terlebih dahulu ketiga soal tersebut ditelaah oleh tiga orang guru, yang bukan pembuat tes tersebut. Selanjutnya dilakukan ujicoba pada ketiga tes tersebut, untuk mendapatkan tes yang memenuhi persyaratan yang baik. Banyaknya butir soal masing-masing tes yang diujicobakan adalah 60 butir soal, dalam bentuk tes objektif dengan lima pilihan jawaban.

Perincian ketiga tes tersebut adalah sebagai berikut: Tes V (tes anchor) terdiri atas 15 butir soal, sedangkan tes I, tes J dan tes K masing-masing 45 butir soal. Berdasarkan analisis hasil ujicoba, banyaknya butir soal yang digunakan sebagai instrumen untuk penyetaraan ditetapkan 40 butir soal (d disesuaikan dengan ketentuan dari Pusat Penelitian dan Pengembangan Departemen Pendidikan Nasional). Ke empat puluh butir soal yang telah dipilih merupakan tes yang sudah *fit* (diterima) pada ketiga tes. Tes yang disetarakan terdiri dari 8 butir soal tes *anchor*, sehingga tes I, tes J dan tes K masing – masing berjumlah 32 butir soal.

Penelitian ini adalah penyetaraan skor tes fisika, maka yang digunakan dalam penelitian adalah skor hasil tes fisika. Alat yang digunakan untuk memperoleh skor tersebut adalah tes prestasi belajar fisika pada Ujian Akhir Nasional (UAN) tahun pelajaran 2003 / 2004 SMA Negeri pada wilayah POKJA 2 Kabupaten Bantul.

Analisis data yang pertama merupakan penelaahan butir soal pada setiap tes. Penelaahan ini dilakukan oleh tiga penelaah secara terpisah. Penentuan diterima/baik tidaknya suatu butir soal, berdasarkan pertimbangan dari ke-16 kriteria yang telah ditetapkan. Jika semua kriteria terpenuhi maka butir soal tersebut diterima/baik. Sebaliknya jika semua kriteria tidak terpenuhi maka butir soal itu dibuang atau ditolak. Butir soal yang tidak memenuhi beberapa atau minimal ada satu kriteria penelaahan perlu direvisi.

Analisis data berikutnya adalah analisis butir soal hasil ujicoba tes, analisis hasil ujicoba tes dan analisis butir tes untuk kepentingan penyetaraan skor tes. Analisis hasil ujicoba dilakukan untuk menentukan karakteristik butir soal

pada setiap tes, dengan bantuan komputer menggunakan program Rascal dari MicroCat versi 3.00. Selanjutnya analisis butir tes untuk penyetaraan dilakukan dengan program Bigsteps MicroCat versi 2.30. Berdasarkan hasil analisis ujicoba, diketahui butir soal berkarakteristik tingkat kesukaran sangat tinggi, menengah dan sangat rendah, dengan melihat parameter tingkat kesukaran butir (b_p). Butir soal yang tingkat kesukarannya terletak pada logis $-2 \leq b_p \leq 2$ termasuk tingkat kesukaran menengah, $b_p > 2$ merupakan butir soal yang sangat sukar, sedangkan $b_p < -2$ merupakan butir soal yang sangat mudah atau membingungkan.

Selanjutnya penentuan suatu butir soal diterima/baik atau revisi ataupun ditolak/dibuang, berdasarkan hasil estimasi parameter butir (tingkat kesukaran) dan kecocokan data terhadap model. Kriteria untuk menentukan butir soal baik atau ditolak/dibuang sebagai berikut:

1. Diterima/baik, $-2 \leq b_p \leq 2$ dan butir soal cocok terhadap model yang digunakan.
2. Ditolak/dibuang, butir soal tidak cocok terhadap model yang digunakan.

Analisis untuk kepentingan penyetaraan skor dilakukan dengan bantuan komputer menggunakan program Bigsteps dari MicroCAT versi 2,30. Analisis yang digunakan untuk membersihkan data, khususnya orang (*testee*) dan untuk memperoleh butir soal yang benar-benar fit. Pembersihan dilakukan untuk, menghindari adanya faktor yang mempengaruhi hasil tes. Faktor – faktor tersebut dapat diketahui melalui data statistik *outfit* untuk setiap *testee*. *Testee* yang memiliki statistik outfit lebih besar dari 2,00 atau ptbis negatif atau keduanya dibuang. Selanjutnya adalah pengkalibrasian butir-butir soal. Butir soal yang mempunyai outfit lebih besar dari 2,00 atau ptbis negatif atau keduanya, harus dibuang. Untuk menentukan persamaan konversi penyetaraan menggunakan Persamaan (11) dan (12). Oleh karena itu diperlukan statistik rerata kemampuan tes *anchor* pada masing-masing tes.

Hasil Penelitian dan Pembahasan

Analisis dilakukan terhadap tiga tes yang akan dipakai perangkat tes I, perangkat tes J dan perangkat tes K. Masing-masing tes dianalisis oleh satu

orang penelaah yang bukan penyusun soal. Penelaah pernah mengikuti penataran tentang Pengelolaan Pengujian, aktif dalam kegiatan MGMP (Musyawarah Guru Mata Pelajaran) dan telah bekerja menjadi guru lebih dari 10 tahun. Sesuai dengan hasil telaah yang dilakukan, hasil analisis dapat dibagi menjadi dua kategori, yaitu butir soal yang diterima tanpa revisi dan butir soal dengan revisi. Butir soal dikategorikan diterima tanpa revisi, apabila semua kriteria yang terdapat pada setiap bidang penelaahan, yaitu meliputi bidang materi, konstruksi dan bahasa terpenuhi.

Butir soal dikategorikan diterima dengan revisi, apabila minimal ada satu kriteria dari ketiga bidang telaah tersebut terpenuhi. Sebaliknya apabila seluruh kriteria tidak terpenuhi pada suatu butir soal, maka tidak diterima. Pada tes I, yang paling banyak direvisi adalah bidang bahasa. Hal ini mengakibatkan tidak terpenuhinya kriteria 14 dan 15, yaitu soal tidak menggunakan bahasa yang komunikatif dan soal tidak menggunakan bahasa yang berlaku setempat. Kemudian bidang konstruksi, tidak terpenuhinya kriteria 4, 5, dan 6, yaitu pokok soal dirumuskan dengan singkat, jelas, dan tegas, pokok soal tidak memberi petunjuk kunci jawaban, pokok soal bebas dari pernyataan yang bersifat negatif ganda. Sedangkan dalam bidang materi, semua kriteria terpenuhi.

Pada tes J yang paling banyak direvisi adalah bidang konstruksi. Hal ini mengakibatkan tidak terpenuhinya kriteria 4, 6, 7 dan 11, yaitu soal dirumuskan dengan singkat, jelas, dan tegas, pokok soal tidak memberi petunjuk kunci jawaban, pokok soal bebas dari pernyataan yang bersifat negatif ganda, gambar/grafik/tabel/diagram dan sejenisnya pada soal jelas dan berfungsi, dan pilihan jawaban yang berbentuk angka disusun berdasarkan urutan besar kecilnya angka. Sedangkan dalam bidang materi tidak terpenuhinya kriteria 1 yaitu soal sesuai indikator dan telaah untuk bidang bahasa mengakibatkan tidak terpenuhinya kriteria 13, yaitu soal menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia.

Pada tes K, yang banyak direvisi adalah bidang telaah materi yang hasilnya tidak terpenuhinya kriteria 1, yaitu soal sesuai indikator. Sedangkan dalam bidang konstruksi kriteria 4 tidak terpenuhi, yaitu pokok butir soal dirumuskan

dengan singkat, jelas dan tegas dan dalam bidang bahasa kriteria 14 tidak terpenuhi, yaitu soal menggunakan bahasa komunikatif.

Tabel 2
Rangkuman butir soal direvisi pada ketiga tes

Tes	Jumlah butir	Persentase	Butir soal yang direvisi
I	8	13,33%	2, 12, 21, 23, 24, 55, 57, 59
J	9	15,00%	4, 5, 8, 10, 23, 24, 34, 35, 56
K	5	8,33%	23, 32, 34, 36, 43

Tabel di atas menunjukkan bahwa soal-soal tes I direvisi sebanyak 8 butir dari 60 butir soal yang ada sebesar 13,33%, sebaliknya 86,67% diterima. Soal-soal tes J yang direvisi sebanyak 9 butir dari 60 butir soal yang ada sebesar 15,00% , sebaliknya 85,00% diterima. Soal-soal tes K yang direvisi sebanyak 5 butir dari 60 butir soal yang ada sebesar 8,33%, sebaliknya 91,67% diterima.

Buti-butir soal pada ketiga perangkat tes yang sudah ditelaah oleh tiga orang penelaah dan beberapa butir soal yang perlu direvisi dimasukkan kembali ke perangkat tes masing-masing, sehingga ketiga perangkat tes dinyatakan telah memenuhi validitas isi. Selanjutnya ketiga perangkat tes akan dilakukan ujicoba untuk mengetahui kualitas butir soal yaitu parameter tingkat kesukaran butir.

Hasil Ujicoba Tes

Ujicoba dilakukan untuk mengetahui kualitas tes/butir soal yang berupa parameter butir (tingkat kesukaran) dan kecocokan butir terhadap model. Ketiga tes diujicobakan pada siswa SMA N 3 Bantul, dan dianalisis dengan bantuan komputer menggunakan program Rascal dari MicroCat versi 3.00. Butir soal yang tingkat kesukarannya terletak pada logis $-2 \leq b_b \leq 2$ termasuk tingkat kesukaran menengah, $b_b > 2$ merupakan butir soal yang sangat sukar, sedangkan $b_b < -2$ merupakan butir soal yang sangat mudah atau membingungkan. Butir soal yang sangat mudah atau sangat sukar dibuang/*deleted*. Butir soal pada tes I dikategorikan dengan tingkat kesukaran menengah,

yaitu sebanyak 48 butir soal. Butir soal lainnya sebanyak 6 butir soal sangat sukar (nomor 5, 22, 36, 46, 51, 53) dan 5 butir soal termasuk sangat mudah (nomor 35, 54, 55, 58, 60) ditolak/dibuang serta 1 soal *deleted*/dibuang (nomor 32). Maka ada 12 butir soal yang tidak memenuhi kriteria parameter butir (tingkat kesukaran) dan 1 butir soal yang dibuang/*deleted*. Butir soal pada tes J, dikategorikan dengan tingkat kesukaran menengah, yaitu sebanyak 45 butir soal. Butir soal lainnya sebanyak 7 butir soal sangat sukar (nomor 5, 6, 32, 36, 46, 47, 51) dan 7 butir soal termasuk sangat mudah (nomor 8, 16, 22, 27, 35, 48, 54) ditolak/dibuang serta 1 soal *deleted*/dibuang (nomor 20). Butir soal pada tes K, dikategorikan dengan tingkat kesukaran menengah, yaitu sebanyak 47 butir soal. Butir soal lainnya sebanyak 7 butir soal sangat sukar (nomor 8, 16, 22, 32, 47, 55, 60) dan 4 butir soal termasuk sangat mudah (nomor 6, 50, 53, 58) ditolak/dibuang serta 2 soal *deleted*/dibuang (nomor 27, 34).

Hasil analisis yang berkaitan antara kecocokan butir dengan model, yaitu model Rasch dengan membandingkan *chi square* hasil perhitungan lebih kecil dari harga *chi square* pada tabel. Derajat kebebasan yang digunakan untuk model satu parameter adalah $m-1$ (m = banyaknya interval kemampuan yang dibutuhkan dalam estimasi, 1 = satu parameter, yakni parameter tingkat kesukaran). Hasil analisis menunjukkan derajat kebebasan untuk tes I, J dan K sebesar 6 atau $m = 7$. Harga derajat kebebasan 6 dan taraf signifikansi yang digunakan 5%, pada tabel diperoleh harga *chi square* 12,59. Apabila harga *chi square* hasil perhitungan lebih kecil dari 12,59 butir soal tersebut secara teori cocok dengan model yang digunakan dan butir soal tersebut diterima. Sebaliknya hasil perhitungan lebih besar dari harga *chi square* tabel maka butir soal tersebut ditolak/dibuang.

Butir-butir soal tes I yang tidak cocok dengan model logistik satu parameter adalah butir soal nomor 27 dan 35, sedangkan butir soal nomor 32 *deleted*. Butir soal lainnya termasuk butir yang cocok dengan model satu parameter(terima). Butir-butir soal tes I yang tidak cocok dengan model logistik satu parameter adalah butir soal nomor 27 dan 35, sedangkan butir soal nomor 32 *deleted*. Butir soal lainnya termasuk butir yang cocok dengan model satu parameter(terima). Butir-butir soal yang tidak cocok terhadap model logistik satu parameter untuk tes J, yaitu butir soal sehingga butir soal nonor 60 tersebut

harus dibuang dan yang *deleted* butir soal nomor 20. Butir soal lainnya termasuk cocok dengan model logistik satu parameter (terima). Butir-butir soal yang tidak cocok terhadap model logistik satu parameter untuk tes K, yaitu butir soal 36 dan yang *deleted* butir soal nomor 27 dan 34. Butir soal lainnya termasuk cocok dengan model logistik satu parameter (terima).

Jumlah butir soal pilihan ganda yang dipakai pada UAN telah ditetapkan dari Dinas P dan K Kabupaten Bantul sebanyak 40 butir. Menurut rangkuman pada Tabel 12 di atas terdapat beberapa soal yang diterima (baik) yang harus dibuang sehingga akan diperoleh jumlah butir soal sebanyak 40 butir soal untuk masing-masing soal. Ketiga butir-butir soal pada tes I, J dan K setiap nomor yang sama memiliki indikator (kisi-kisi) yang sama. Butir-butir soal yang dipakai untuk kepentingan tes UAN sekaligus penyetaraan skor diambil dari butir-butir soal tes I, J dan K yang diterima (baik) dengan nomor-nomor butir soal yang sama pada ketiga tes, sehingga tes I, J dan K mempunyai 40 butir soal pilihan ganda dengan indikator yang sama. Di antara 40 butir soal pada masing-masing tes terdapat 8 butir soal *anchor* yaitu nomor 7, 19, 21, 30, 37, 38, 52, dan 56.

Menurut Saifuddin Azwar (1996 : 135-136) butir yang dibuang, apabila suatu butir soal sedemikian sukarnya sehingga tidak seorang siswapun yang dapat menjawabnya dengan benar dan suatu butir soal yang sedemikian mudahnya sehingga seluruh siswa dapat menjawab dengan benar, maka butir tersebut tidak akan banyak gunanya dalam membedakan antara mereka yang menguasai bahan pelajaran dan mereka yang tidak.

Berdasarkan uraian di atas, terdapat 20 butir soal yang dibuang. Dibuangnya beberapa butir soal tidak mempengaruhi validitas isi dari ketiga tes tersebut. Butir-butir soal yang dibuang sudah diwakili oleh butir soal lain pada masing-masing subpokok bahasan dari materi fisika yang tertuang pada standar kompetensi tes. Apabila butir soal pada pilihan ganda belum sesuai dengan Standar Kompetensi Lulusan dan Spesifikasi Ujian Akhir SMA/MA yang disusun oleh Pusat Penilaian Pendidikan Badan Penelitian dan Pengembangan Departemen Pendidikan Nasional Tahun 2004, maka butir soal dibuat untuk materi soal uraian.

Penyetaraan Skor Tes

Hasil analisis ujicoba terhadap tes I, J, dan K terdapat beberapa butir soal yang dibuang dan disusun tes baru masing-masing sebanyak 40 butir soal. Tes I terdiri dari tes I dan tes V, tes J terdiri dari tes J dan tes V serta tes K terdiri dari tes K dan tes V. Tes V merupakan tes *anchor* yang sama pada ketiga tes, sedangkan tes I, tes J dan tes K merupakan tes yang disetarakan.

Tes hasil analisis yang telah disempurnakan kemudian diteskan pada tiga kelompok siswa, yaitu tes I diberikan pada kelompok A, tes J diberikan pada kelompok B dan tes K diberikan pada kelompok siswa C. Tes *anchor* (V) diberikan pada tiga kelompok siswa tersebut. Hasil jawaban siswa dianalisis menggunakan program Bigsteps maka akan diperoleh data tingkat kesukaran dan kemampuan *testee* pada ketiga tes untuk keperluan penyetaraan. Hasil analisis ketiga soal dengan menggunakan program Bigsteps dari MicroCAT versi 2,30 terlampir.

Metode penyetaraan yang digunakan adalah rerata dan sigma yang sifatnya simetri, maka penyetaraan dapat dilakukan antartes mana saja, seperti penyetaraan skor tes I dan tes J ke tes K, tes K sebagai standar atau penyetaraan tes I dan tes K ke tes J, dengan tes J sebagai standar atau penyetaraan tes J dan tes K ke tes I dengan tes I sebagai standar. Hasil penyetaraan yang diperoleh pada prinsipnya tidak berbeda. Dalam penelitian ini dilakukan penyetaraan skor tes I dan tes J ke tes K, tes K sebagai standar dan penyetaraan skor tes I dan tes K ke tes J, tes J sebagai standar.

Penyetaraan Tingkat Kesukaran Butir Soal

Butir soal *anchor* yang dimasukkan ke dalam masing-masing tes bersifat internal. Butir-butir pada tes *anchor* diseleksi dengan cara yang sama, nomor butir soal yang dibuang pada satu tes akan mengakibatkan ditolaknya pada tes yang lain. Dengan demikian butir tes *anchor* pada ketiga tes mempunyai nomor butir yang sama dan jumlahnya juga sama. Misalnya butir tes *anchor* nomor 3 pada tes I sama dengan butir soal pada tes J dan juga sama dengan butir soal pada nomor 3 tes K. Pada penelitian ini banyaknya tes *anchor* yang digunakan berjumlah 8 butir soal, sehingga jumlah butir soal *anchor* mempunyai jumlah

20% dari jumlah total pada masing-masing tes. Hasil seleksi tes *anchor* yang memenuhi kriteria menurut tingkat kesukaran butir soal disajikan dalam tabel berikut:

Tabel 3
Tingkat kesukaran butir tes anchor pada tes I, tes J, dan tes K

No. butir	b_{vi}	b_{vj}	b_{vk}
23	-0.36	-0.58	-0.02
26	0.18	0.42	-0.69
27	0.46	0.76	0.16
37	-0.26	-0.80	-1.10
38	0.13	-0.48	-0.24
Σ	0.15	-0.68	-1.89
μ	0.03	-0.136	-0.378

Tabel di atas memperlihatkan rerata tingkat kesukaran butir tes *anchor* yang digunakan dalam melakukan penyetaraan tingkat kesukaran antartes. Penyetaraan dilakukan dengan menggunakan Persamaan 14. Penyetaraan dilakukan pada butir fit dan pada ketiga tes diambil nomor butir yang sama-sama fit untuk menentukan konversi tingkat kesukaran butir antartes diawali dengan menentukan nilai konstanta b (untuk menentukan b digunakan Persamaan 8 dengan $a = 1$), selanjutnya ditentukan persamaan konversi tingkat kesukaran butir. Perhitungan untuk menentukan persamaan konversi tingkat kesukaran butir tes I dan tes J ke tes K secara langsung dapat dilihat pada lampiran di belakang. Persamaan konversi tingkat kesukaran butir ditunjukkan pada tabel berikut:

Tabel 4
Persamaan konversi tingkat kesukaran butir

No.	Kesetaraan	Persamaan konversi
1.	Tes I dengan tes K	$b_{KI}^* = b_I - 0.408$
2.	Tes J dengan tes K	$b_{KJ}^* = b_J - 0.242$

Keterangan:

b_I = tingkat kesukaran awal tes I

b_J = tingkat kesukaran awal tes J

b_{KI}^* = konversi tingkat kesukaran tes I ke tes K

b_{KJ}^* = konversi tingkat kesukaran tes J ke tes K

Berdasarkan Tabel 15, diperoleh hasil konversi tingkat kesukaran butir dari tes I ke tes K dan tes J ke tes K.

Penyetaraan Kemampuan *Testee*

Dengan mensubstitusikan konstan translasi b pada tingkat kesukaran butir soal, dapat ditentukan persamaan kemampuan *testee*. Perhitungan persamaan konversi kemampuan *testee* dapat dilihat pada lampiran.

Penyetaraan kemampuan *testee* dari tes I dan tes J ke tes K sebagai standar. Persamaan konversi yang diperoleh ditunjukkan pada Tabel 5.

Tabel 5
Persamaan konversi kemampuan testee

No.	Kesetaraan	Persamaan konversi
1.	Tes I dengan tes K	$\theta_{KI}^* = \theta_I - 0.408$
2.	Tes J dengan tes K	$\theta_{KJ}^* = \theta_J - 0.242$

Keterangan :

θ_I = kemampuan awal *testee* pada I

θ_J = kemampuan awal *testee* pada J

θ_{KI}^* = konversi kemampuan *testee* untuk tes I ke tes K

θ_{KJ}^* = konversi kemampuan *testee* untuk tes J ke tes K

Berdasarkan uraian yang telah dikemukakan terdahulu melalui penyetaraan dapat dilakukan perbandingan tingkat kesukaran butir antartes atau kemampuan *testee* yang menempuh tes yang berbeda.

Simpulan

Berdasarkan hasil penelitian, dapat disampaikan beberapa simpulan. Perbandingan nilai parameter tingkat kesukaran butir ketiga tes adalah bahwa dari ketiga perangkat tes tersebut urutan tingkat kesukaran butir dari yang paling sukar adalah: perangkat tes K, perangkat tes J, dan perangkat tes I. Persamaan konversi untuk penyetaraan ketiga tes sebagai berikut.

Persamaan konversi untuk tingkat kesukaran butir :

$$- b_{KI}^* = b_I - 0,408$$

$$- b_{KJ}^* = b_J - 0,242$$

Menurut kedua persamaan konversi untuk tingkat kesukaran butir di atas, maka tingkat kesukaran butir tes K lebih sukar dibanding tingkat kesukaran butir tes I dan tes J, sedangkan tingkat kesukaran butir tes J lebih sukar dari tes I.

Persamaan konversi untuk kemampuan *testee*:

$$- \theta_{KI}^* = \theta_I - 0,408$$

$$- \theta_{KJ}^* = \theta_J - 0,242$$

Menurut kedua persamaan konversi untuk kemampuan *testee* di atas, maka kemampuan *testee* penempuh tes K lebih tinggi dibanding kemampuan *testee* penempuh tes I dan tes J, sedangkan kemampuan *testee* penempuh tes J lebih tinggi dari penempuh tes I.

Berdasarkan kesimpulan di atas, dapat dikemukakan beberapa hal mengenai perlunya penyetaraan tes:

1. Ujian yang menggunakan beberapa perangkat soal yang sudah diketahui tingkat kesukaran butir soal dapat digunakan beberapa sekolah yang mempunyai mutu berbeda, misalnya untuk sekolah yang mutunya baik menggunakan soal yang tingkat kesukaran butir lebih tinggi dibanding sekolah yang mempunyai mutunya kurang baik.
2. Soal-soal yang dikalibrasi ke dalam suatu skala dengan model Rasch dan cocok dengan model dapat disimpan dalam suatu bank soal yang

dikomputerisasi. Soal-soal yang hendak disimpan dalam suatu bank soal, agar mempunyai skala yang sama diperlukan penyetaraan antar paket-paket soal. Beberapa paket-paket soal yang sudah dikalibrasi dan diketahui tingkat kesukaran butirnya dapat untuk memperkaya penyusunan bank soal.

Daftar Pustaka

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole.
- Bahrul Hayat (1996). *Pengantar model Rasch*. Makalah disampaikan pada Pelatihan Standarisasi Tes Prestasi Belajar di Cisarua Bogor, Jawa Barat. Jakarta: PUSISJIAN Balitbang Depdikbud.
- Bahrul Hayat (1997). *Analisis butir soal dengan bigsteps*. Jakarta: Pusat Penelitian dan Pengembangan Sistem Pengujian Balitbang Depdikbud.
- Cronbach, L. J. (1984). *Essentials of psychology testing*. New York: Haper and Row, Publisher, Inc.
- Dali S. Naga (1992). *Pengantar teori skor pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Depdikbud. (1993). *Kurikulum sekolah menengah umum*. Jakarta: Dirjen. Dikmenum
- Depdikbud. (1994). *Bahan penataran pengujian*. Jakarta: Pusat Penelitian dan Pengembangan sistem Pengujian.
- Depdiknas. (2000). *Pengembangan Silabus dan Penilaian*. Jakarta: Dirjen Dikdasmen
- Djemari Mardapi (1991). Konsep dasar teori respon butir. Perkembangan dalam bidang pengukuran pendidikan. *Cakrawala Pendidikan* 3(X)1-16.
- Djemari Mardapi (1997). *Ragam bentuk evaluasi*. Bahan Semiloka Evaluasi Sistem Penilaian dan pengukuran Prestasi belajar Mahasiswa UGM. Yogyakarta.
- Djemari Mardapi (1999). Estimasi kesalahan pengukuran dalam bidang pendidikan dan implikasinya pada ujian nasional. *Pidato pengukuhan Guru Besar*. Disampaikan di depan Rapat Senat Terbuka UNY, 11 September 1999.
- Erna Miyatun (2000). *Komparasi metode penyetaraan tes menurut teori respon butir*. Tesis: Universitas Negeri Yogyakarta.

- Gronlund, N. E., (1976). *Measurement and evaluation in teaching*. New York: Macmillan.
- Guenther, R. D., (1990). *Modern optics*. New York: John Wiley & Sons.
- Hariwibowo (1994). *Mengupayakan hasil ujian menjadi akurat dan obyektif*. Buletin Pengujian dan Penelitian. Jakarta: Pusat Penelitian dan pengembangan system Pengujian Balitbang Depdikbud.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer Academic Publiser.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J., (1991). *Fundamentals of item response theory*. New Delhi: Sage Publication, Inc.
- Kerlinger, F. N. (1986). *Asas-asas penelitian behavioral* (Terjemahan L.R. Simatupang). Yogyakarta: Gadjah Mada University Pres.
- Kittel, C. (1996). *Introduction to solid state physics*(7th ed). New York: John Wiley & Sons, Inc.
- Klein, L. W., & Jarjoura, D. (1985). The importance of content representation for common item equating with non random group. *Journal of Educational Measurement* .22, 197 – 206.
- Kolen, M. J., & Brennan, R. L. (1995) *Test equating: Methods and practices*. New York: Springer.
- Nonny Swediati (1998). *Metode untuk penyetaraan (equating) skor tes secara klasik*. Jakarta: Pusat Pengujian, Balitbang Depdikbud.
- Saifuddin Azwar. (1996). *Tes prestasi: Fungsi dan pengembangan pengukuran prestasi belajar (Edisi ke-2)*. Yogyakarta: Pustaka Pelajar
- Suke Silverius (1991). *Evaluasi hasil belajar dan umpan balik*. Jakarta: Grasindo
- Sumadi Suryabarata (1987). *Pengembangan tes hasil belajar*. Jakarta: Rajawali Pers
- Sumadi Suryabarata (1990). *Psikologi Pendidikan*. Jakarta: Rajawali Pers.
- Tim Sisjian (1995). *Bank Soal* . Jakarta: Pusat Penelitian dan Pengembangan sistem Pengujian
- White, H. E., (2001). *Fundamentals of optics* (4th ed). Berkeley: University of California.
- Wright, B. D., & Stone, M. H., (1979). *Best test design*. Chicago: MESA Press.
- Yahya Umar. (1991). *Menuju pembaharuan sistem pengujian pendidikan*. Jakarta: Pusiasjian Balitbang Depdikbud.
- Young, H. D., & Freedman R. A., (2000) *University physics* (10th ed). Santa Barbara: Addison Wesley Longman, Inc.