

## **PENYUSUNAN PERANGKAT SOAL UJIAN AKHIR MATA PELAJARAN SAINS–BIOLOGI SMP DALAM RANGKA PENGEMBANGAN BANK SOAL**

*Oleh:*  
*Kustriyono*

### **Abstrak**

Penelitian ini bertujuan untuk memperoleh butir soal Ujian Akhir mata pelajaran Sains-Biologi SMP yang memenuhi syarat untuk pengembangan bank soal. Model Rash digunakan untuk menentukan parameter butir.

Peserta uji coba adalah kelas 3 SMP, baik negeri maupun swasta. Penentuan SMP ini didasarkan pada mutu sekolah dengan kategori baik, sedang, dan kurang. Paket soal yang diujikan meliputi dua paket soal yang setiap paketnya berisi 50 butir soal. Soal paket A dikerjakan oleh 771 siswa, sedangkan soal paket B dikerjakan oleh 759 siswa sebagai peserta uji coba. Analisis butir soal dilakukan melalui analisis kualitatif dan analisis kuantitatif. Analisis kualitatif didasarkan pada kesesuaian butir soal dengan kartu telaah butir dengan 16 kriteria dari segi materi, konstruksi, dan bahasa. Analisis kuantitatif dilakukan melalui dua program analisis butir soal yaitu program Iteman untuk analisis dengan Teori Tes Klasik dan program Bigsteps untuk analisis dengan Teori Respons Butir model satu parameter.

Hasil analisis menunjukkan bahwa dari 50 butir soal paket A diperoleh 47 butir soal yang relevan untuk diangkat ke bank soal, sedangkan dari paket B diperoleh 46 butir soal. Rerata tingkat kesukaran soal paket A sebesar 0,45 logit, sedangkan soal paket B sebesar 0,32 logit. Besarnya daya pembeda soal paket A antara 0,07 sampai 0,48 logit sedang soal paket B antara 0,08 sampai 0,45 logit. Untuk soal paket A, nilai fungsi informasi maksimum berada pada  $q=0,0$  sedang soal paket B pada  $q=0,5$ . Soal paket A lebih efektif 1,049 kali dari pada soal paket B. Formula konversi sekor untuk penyetaraan paket soal A ke soal paket B adalah  $B = 0,659949 a - 0,226549$ . Soal paket A lebih sukar dibanding soal paket B. Ke-93 butir soal yang meliputi soal paket A dan soal paket B dapat diangkat ke bank soal dalam rangka pengembangan bank soal di Kabupaten Wonosobo.

**Kata Kunci:** *soal ujian akhir sains, analisis butir, bank soal.*

## **Pendahuluan**

Pendidikan merupakan usaha yang disengaja dengan tujuan agar peserta didik berkembang melalui proses pembelajaran. Keberhasilan proses pembelajaran ditentukan oleh banyak faktor, di antaranya faktor siswa, materi, media belajar, guru, lingkungan, dan lain-lain. Faktor guru yang turut mempengaruhi keberhasilan dalam proses pembelajaran antara lain kemampuan menguasai bahan pelajaran, mengelola kelas, kemampuan menggunakan media dan sumber belajar, serta kemampuan mengevaluasi hasil belajar siswa.

Penilaian merupakan komponen yang penting dalam sistem pendidikan karena mencerminkan perkembangan atau kemajuan hasil pendidikan dari satu waktu ke waktu lain. Di samping itu, berdasarkan penilaian, tingkat ketercapaian prestasi pendidikan antara satu sekolah dengan sekolah lain atau satu wilayah dengan wilayah lain dapat dibandingkan (Depdiknas, 2000: 7).

Dalam rangka peningkatan keterampilan membuat soal yang baik, seorang guru diharuskan melakukan analisis butir soal yang pernah dibuat, sehingga akan diketahui butir soal mana yang baik dan mana yang harus diperbaiki atau diganti. Namun, kenyataannya banyak guru hanya melakukan analisis yang bersifat global, yaitu hanya dilihat dari skor secara keseluruhan, belum sampai pada tingkat kesulitan, daya beda, serta efektivitas distraktor tiap-tiap butir soal. Dengan demikian, kualitas soal buatan guru, baik yang berupa ulangan harian, ulangan semester, serta ujian akhir masih perlu dipertanyakan.

Dengan diberlakukannya otonomi pendidikan, berarti pemerintah memberikan kewenangan kepada daerah untuk menentukan kebijakan-kebijakan yang berkaitan dengan masalah pendidikan. Sebelumnya, mata pelajaran yang diujikan secara nasional pada jenjang SMP meliputi 6 mata pelajaran (dulu bernama EBANAS) kini hanya tinggal 3 mata pelajaran saja, yaitu Bahasa Indonesia, Matematika, dan Bahasa Inggris. Untuk mata pelajaran IPA termasuk di dalamnya Biologi diserahkan kepada daerah, dalam hal ini Diknas kabupaten/kota agar menyusun sendiri perangkat untuk ujian akhir (UNAS). Dengan demikian, Diknas kabupaten/kota diberi kewenangan untuk membuat perangkat ujian dalam hal ini butir-butir soal Sains-Biologi dan diberikan hak untuk memutus siswa yang dianggap berhasil (lulus) dan yang

tidak berhasil berdasarkan ketentuan minimal standar kompetensi yang harus dikuasai siswa.

Ada dua pendekatan yang digunakan untuk mengetahui kualitas tes yang dibuat, yaitu pendekatan kualitatif dan kuantitatif. Pendekatan kualitatif dilakukan dengan cara menelaah butir soal, hal ini dilakukan sebelum perangkat tes tersebut diujikan. Pendekatan kualitatif pada dasarnya menekankan pada penilaian dari aspek materi, konstruksi dan bahasa. Depdiknas melalui Puspendik telah membuat pedoman penelaahan terhadap butir tes, baik dalam bentuk uraian maupun dalam bentuk pilihan ganda. Butir tes yang secara kualitatif sudah baik belum tentu baik pula secara empiris.

Pendekatan kuantitatif merupakan metode analisis butir tes yang didasarkan pada data empiris yang diperoleh melalui respons peserta tes. Dengan demikian, pendekatan kuantitatif dilakukan setelah tes diberikan kepada peserta tes atau pernah diujikan pada kelompok peserta lain yang memiliki karakteristik hampir identik.

Salah satu langkah untuk menunjang kelembagaan dan profesionalisasi dalam penilaian pendidikan adalah penyiapan alat ukur yang valid dan reliabel. Untuk mendapatkan alat ukur yang valid dan reliabel tersebut diperlukan langkah-langkah dan tahapan-tahapan pengembangan yang baku (standar). Soal yang telah diperiksa melalui tahapan-tahapan tersebut kemudian disimpan secara sistematis di dalam suatu tempat yang disebut "bank soal". Bank soal akan mendukung penilaian pendidikan yang valid dan reliabel (Depdiknas, 2000: 9).

Berdasarkan pembatasan masalah yang telah dikemukakan di atas, rumusan masalah yang peneliti ajukan adalah sebagai berikut.

1. Bagaimana langkah validasi paket soal Ujian Akhir mata pelajaran Sains-Biologi SMP agar layak diangkat untuk pengembangan bank soal di Kabupaten Wonosobo?
2. Bagaimana karakteristik butir-butir soal yang merupakan indikator butir soal Ujian Akhir mata pelajaran Sains-Biologi SMP?
3. Berdasarkan hasil analisis empiris berapa banyak butir soal ujian akhir mata pelajaran Sains-Biologi SMP yang dapat disimpan ke dalam bank soal di Kabupaten Wonosobo?

Berdasarkan rumusan masalah di atas, tujuan penelitian yang hendak dicapai sebagai berikut.

1. Memberi pedoman pembuatan perangkat soal Ujian Akhir mata pelajaran Sains-Biologi SMP yang butir-butir penyusunnya mempunyai karakteristik bank soal di Kabupaten Wonosobo.
2. Memperoleh harga statistik butir soal yang merupakan indikator ciri-ciri butir soal Ujian Akhir mata pelajaran Sains-Biologi SMP.
3. Memperoleh 2 perangkat soal Ujian Akhir mata pelajaran Sains-Biologi SMP yang butir-butir penyusunnya memiliki karakteristik butir soal yang baik.

Menurut Azwar (2001: 2), dilihat dari wujud fisiknya, suatu tes tidak lain daripada sekumpulan pertanyaan yang harus dijawab dan/atau tugas yang harus dikerjakan yang memberikan informasi mengenai aspek psikologis tertentu berdasarkan jawaban terhadap pertanyaan-pertanyaan atau cara dan hasil subjek dalam melakukan tugas-tugas tersebut. Jadi, tes terdiri dari sejumlah pertanyaan yang memiliki jawaban betul atau salah. Sesuai dengan bentuk, soal betul salah ini bisa seluruhnya atau sebagian.

Muhamad Noer (1987: 15) mengemukakan langkah-langkah yang harus diikuti dalam menyusun suatu tes, yaitu: (1) menentukan tujuan utama penggunaan skor tes, (2) menentukan konstruk yang hendak diukur atau menentukan domain, (3) menyiapkan spesifikasi tes, menetapkan proporsi butir yang harus berpusat pada setiap jenis kriteria, (4) menyusun pool awal butir, (5) mengadakan penelaahan kembali terhadap butir-butir yang telah dibuat dan melakukan revisi bila diperlukan, (6) melakukan uji coba pendahuluan dan merevisi bila diperlukan, (7) melaksanakan uji lapangan, (8) menentukan ciri-ciri statistik skor butir, (9) merencanakan dan melaksanakan pengkajian validitas dan reliabilitas untuk bentuk akhir tes, dan (10) mengembangkan panduan administrasi penskoran dan penafsiran skor tes.

Performansi subjek pada suatu skala pengukuran atau test psikologis dinyatakan dalam bentuk skor (*score*). Skor tidak lain adalah harga suatu jawaban terhadap pertanyaan dalam tes. Skor ini merupakan skor perolehan (*obtained score* atau *observed score*) yang sering disebut skor tampak. Di samping itu, selain

skor tampak, ada pula skor lain yang merupakan skor sesungguhnya. Skor sesungguhnya merupakan angka performansi yang benar dan murni dan tidak pernah dapat diketahui besarnya karena tidak dapat diungkap secara langsung oleh tes. Skor sesungguhnya ini sering disebut skor murni (*true score*). Selain skor tampak dan skor murni, dalam setiap pengukuran selalu terdapat kesalahan (*error*) yang besarnya tidak dapat diketahui. Teori skor murni menjelaskan bagaimana kesalahan pengukuran bisa mempengaruhi skor tampak.

Tes yang baik harus memberikan bukti akan keunggulan. Untuk memperoleh tes yang baik harus dimulai dengan perencanaan yang baik sampai uji coba tes. Hasil uji coba merupakan masukan untuk perbaikan butir tes sehingga menjadi butir tes yang baik. Masukan tersebut merupakan gambaran karakteristik suatu tes yang meliputi: distribusi respons, distribusi indeks kesukaran, distribusi daya pembeda, kehandalan, kesahihan prediktif, dan kesahihan konstruk yang dinyatakan dengan dimensi tes. Semua indeks yang menyatakan katakarakteristik suatu tes dapat dicari dengan menggunakan teori tes klasik (Allen & Yen, 1979: 72).

Pada saat ini ada dua teori pengukuran yang berkembang dan banyak digunakan dalam merancang dan menganalisis alat ukur atau tes. Pertama adalah teori tes klasik yang dikembangkan sejak tahun 1940 dan telah digunakan secara luas, sedang teori yang kedua adalah teori respons butir, muncul setelah teknologi komputer berkembang. Teori yang kedua ini lebih banyak menggunakan asumsi dibandingkan dengan teori yang pertama, namun dapat menyajikan informasi lebih banyak (Mardapi, 1998: 25).

Sejalan dengan itu, Depdiknas (2000: 53) menyebutkan bahwa dalam melakukan analisis kuantitatif, dikenal dua cara yaitu analisis cara tradisional dengan menggunakan teori tes klasik dan analisis modern dengan menggunakan *Item Response Theory*. Pada dasarnya, ada dua macam karakteristik soal ditinjau dari analisis soal secara empirik, yaitu tingkat kesukaran dan daya pembeda soal. Daya pembeda soal adalah kemampuan suatu soal dalam membedakan antara siswa-siswa yang mampu (menguasai materi) dengan siswa-siswa yang kurang mampu (kurang menguasai materi). Pendekatan yang digunakan untuk mendapatkan karakteristik soal dengan cara tradisional/klasikal berbeda dengan

cara modern. Makin tinggi indeks daya pembeda, berarti makin tinggi kemampuan soal tersebut dalam membedakan siswa yang menguasai materi dan yang kurang menguasai materi yang dinyatakan oleh soal tersebut. Khusus untuk soal pilihan ganda, karakteristik lain yang perlu diperhatikan adalah sejauh mana pilihan jawaban (*option*) berfungsi sebagai pengecoh.

Tiga asumsi penting yang digunakan pada teori tes klasik adalah: pertama, tidak ada korelasi antara skor yang sebenarnya dan skor kesalahan; kedua, tidak ada korelasi antara kesalahan acak pada pengukuran yang dilakukan berulang-ulang; dan ketiga, rerata kesalahan acak pengukuran sama dengan nol. Dengan menggunakan tiga asumsi tersebut, selanjutnya dikembangkan formula-formula untuk menghitung besarnya indeks kesahihan dan indeks kehandalan suatu alat ukur. Tingkat kesukaran dan daya pembeda butir juga dikembangkan untuk menganalisis suatu tes.

Menurut Azwar (2001: 131), pada analisis butir dengan menggunakan teori tes klasik tipe objektif, kualitas butir dilihat dari paling tidak dua parameter atau dua kriteria, yaitu (1) indeks kesukaran butir, dan (2) indeks daya diskriminasi atau indeks daya pembeda butir. Lebih lanjut, Azwar mengatakan, analisis butir akan menguji pula efektivitas distraktor-distraktor pada setiap butir untuk menentukan apakah distraktor tersebut sudah berfungsi atau belum.

Teori tes modern yang dikenal dengan nama teori respons butir (*item response theory*) mencoba mengatasi kelemahan-kelemahan yang dimiliki teori tes klasik. Teori ini dimunculkan oleh Lord dalam disertasinya pada tahun 1952 yang dikenal dengan teori skor tes. Selanjutnya, Birnbaum mengembangkan dasar statistik untuk model teori respons butir pada tahun 1957. Pada tahun 1960 Rasch, seorang matematikawan Swedia mengembangkan model probabilistik untuk tes intelegensi dan tes pencapaian belajar. Model ini dikenal dengan Model Rasch yang selanjutnya dikembangkan di Amerika mulai tahun 1967 oleh Benyamin Wright dan pengikutnya di Universitas Chicago (Mardapi, 1991: 3).

Teori respons butir mengenal tiga macam model logistik. Perbedaan ketiga model tersebut terletak pada perbedaan banyaknya parameter yang dipakai untuk menggambarkan karakteristik butir dalam model yang bersangkutan. Dalam teori respons butir dikenal tiga parameter butir yaitu: (1)

indeks kesukaran butir, (2) daya pembeda butir, dan (3) probabilitas tebakan semu.

Model 1P (satu parameter) dikenal dengan model Rasch. Model Rasch adalah suatu formulasi matematik yang menghubungkan kemungkinan (*probability*) hasil (benar atau salah untuk butir yang dikotomi) apabila seseorang menjawab suatu butir soal sesuai dengan karakteristik orang dan butir soal tersebut. Apabila kemampuan orang lebih tinggi dari tingkat kesukaran soal nilai kemungkinan orang tersebut menjawab soal dengan benar lebih besar dari 0,50, begitu juga sebaliknya, apabila kemampuan orang lebih rendah dari tingkat kesukaran soal nilai kemungkinan orang tersebut menjawab benar kurang dari 0,50. Jika kemampuan orang sama persis dengan tingkat kesukaran soal nilai kemungkinan orang tersebut menjawab soal dengan benar adalah 0,05 (Depdikbud, 1999: 143).

Formula yang digunakan pada model 1-P adalah:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

Keterangan:

- $P_j$  = probabilitas jawaban benar butir ke-j
- $q$  = kemampuan
- $D$  = 1,7
- $e$  = 2,71828
- $b_j$  = tingkat kesukaran butir ke-j

Pada teori respons butir dikenal fungsi informasi yang digunakan untuk menggambarkan kekuatan butir soal atau suatu paket soal (Hambleton dan Swaminathan, 1991: 94). Fungsi informasi menyatakan kekuatan atau sumbangan butir soal dalam mengungkapkan *latent trait* yang diukur dengan soal tersebut. Dengan fungsi informasi dapat diketahui butir mana yang memberi informasi paling benar sehingga membantu dalam seleksi butir soal (Miyatun, 2000).

Pada model logistik 1-P, fungsi informasi butir memenuhi persamaan sebagai berikut.

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

**Keterangan:**

D = konstanta, 1,7

q = skala kemampuan

e = 2,71828

$b_i$  = tingkat kesukaran butir ke-I

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut (Hambleton dan Swaminathan, 1991: 94) sehingga fungsi informasi paket tes akan tinggi jika butir penyusunnya mempunyai fungsi informasi yang tinggi pula. Secara matematis fungsi informasi tes dapat ditulis:

$$I_i(\theta) = \sum_{i=1}^n I_i(\theta)$$

Nilai-nilai parameter butir dan peserta merupakan hasil estimasi sehingga kebenarannya bersifat probabilistik dan tidak terlepas dari kesalahan pengukuran. Dalam teori respons butir, kesalahan tersebut disebut kesalahan baku pengukuran (*standard error of measurement. Standard error of measurement (SEM)*) berkaitan erat dengan fungsi informasi. Fungsi informasi dan SEM merupakan hubungan yang berbanding terbalik kuadratik. Makin besar nilai fungsi informasi berarti SEM makin kecil dan sebaliknya. Jika fungsi informasi dinyatakan dengan  $I(q)$  dan kesalahan baku pengukuran dinyatakan dengan  $SEM(q)$ , bentuk hubungan keduanya dapat dirumuskan sebagai berikut.

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

Keunggulan teori tes klasik terletak pada kemudahan dalam pemahaman konsep dan penggunaannya sehingga banyak digunakan. Kelemahan teori ini terletak pada hasil estimasi parameter yang tergantung pada karakteristik peserta tes dan hasil estimasi parameter kemampuan tergantung pada karakteristik butir. Tes yang mudah akan menghasilkan estimasi kemampuan yang tinggi.

Demikian pula sebaliknya, tes yang sulit akan menghasilkan estimasi kemampuan yang rendah.

Menurut teori tes klasik, besarnya tingkat kesukaran butir yang baik adalah 0,30 sampai 0,70 (Allen & Yen, 1979: 121), sedangkan daya pembeda yang diterima minimum 0,20. Pengecoh dikatakan berfungsi bila terdapat peserta yang memilih tiap pengecoh. Indeks kehandalan tes yang baik menurut Linn (Mardapi, 1998) adalah minimum 0,70.

Menurut Depdiknas (2000: 10) soal yang diterima atau soal yang baik menurut teori respons butir harus memenuhi kriteria sebagai berikut.

- 1) Statistik daya pembeda soal tidak negatif.
- 2) Statistik OUTFIT soal tidak lebih besar dari 2,00.
- 3) Distribusi jawaban logis.

Mengenai patokan standar butir soal yang diterima dalam rangka pengembangan bank soal, Hambleton, Swaminathan, dan Rogers (1991: 54) mengemukakan bahwa: "*item difficulty value were chosen to be similar to those commonly found in practise (-2.00 to +2.00)*". Rasmuin (2002) menyebutkan bahwa butir dikatakan tidak cocok dengan model Rasch apabila *oufit* > 2,00 logit dan *ptbis* < 0,00 logit. Syahrial menambahkan, bahwa butir soal tidak cocok dengan model jika *oufit* > 2,00 atau korelasi point biserial bernilai negatif. Tingkat kesukaran ideal membentang dari -2,00 sampai dengan 2,00. Indeks reliabilitas diketahui dari harga fungsi informasi minimal 0,70.

Salah satu langkah untuk menunjang kelembagaan dan profesionalisasi dalam penilaian pendidikan adalah penyiapan alat ukur yang valid dan reliabel. Untuk mendapatkan alat ukur yang valid dan reliabel tersebut diperlukan langkah-langkah dan tahapan-tahapan pengembangan yang baku (standar). Soal yang telah lulus melalui tahapan-tahapan tersebut kemudian disimpan secara sistematis di dalam suatu tempat yang disebut sebagai "Bank Soal". Adanya bank soal akan mendukung penilaian pendidikan yang valid dan reliabel (Depdiknas, 2000: 1).

Pada praktiknya, bank butir dapat didefinisikan secara berbeda tergantung dari tujuan penggunaannya. Menurut Choppin (Umar, 1999: 207), bank butir adalah koleksi butir-butir soal yang terorganisasi dan sudah melalui proses pengukuran karakteristik butir (tingkat kesukaran, reliabilitas, validitas, dll).

Umar (1999: 209) menyatakan alasan perlunya bank butir antara lain.

- 1) Adanya kebijakan desentralisasi ujian nasional menuntut agar hasil ujian tetap bermutu dan dapat dikomparasikan.
- 2) Menyangkut masalah biaya dan waktu, bila konstruksi soal ujian harus dibuat setiap menyelenggarakan ujian.
- 3) Karena besarnya jumlah butir dalam bank butir maka tingkat kebocoran soal sangat kecil.
- 4) Menyangkut masalah kualitas soal ujian dimana butir-butir soal yang baik dalam bank butir telah diketahui ciri-cirinya sehingga konstruksi soal ujian yang berkualitas dapat terpenuhi.
- 5) Guru dapat menyusun perangkat ujian dengan melakukan *sharing* terhadap bank butir.
- 6) Melalui bank butir, penyusunan tes yang baik untuk kepentingan yang berbeda sangat mungkin untuk dilakukan.
- 7) Guru dapat berkonsentrasi pada penyampaian pelajaran tanpa harus menyisihkan waktu untuk menyusun tes ketika akan ada ujian.
- 8) Meski mahal, bank butir dapat menjadi akses pendidikan yang menyeluruh (*comprehensive*).
- 9) Proses kalibrasi menghasilkan sistem yang spesifik tentang subjek sehingga penyusunan peta kurikulum dan strategi guru dapat direncanakan.

Dalam mengembangkan bank soal sebagai langkah pembakuan tes dan butir soal, perlu dilakukan tahapan-tahapan antara lain: (1) penyusunan kisi-kisi, (2) penulisan soal, (3) telaah (review) dan perbaikan (revisi) soal, (4) perakitan tes, (5) uji coba tes, (6) entri data, (7) analisis butir soal, (8) seleksi soal, dan (9) pengkartuan/komputerisasi bank soal (Depdiknas, 2000: 2).

## **Metode Penelitian**

Penelitian ini dilaksanakan di SMP Kabupaten Wonosobo, meliputi SMP Negeri dan SMP Swasta yang berperan sebagai responden uji coba soal tes yang akan dimasukkan ke dalam bank soal IPA-Biologi di Kabupaten Wonosobo. Penelitian ini dilaksanakan pada akhir tahun pelajaran 2003/2004, dimulai pada awal bulan April sampai dengan akhir bulan Mei 2004.

Penelitian ini termasuk penelitian deskriptif evaluatif. Penelitian deskriptif evaluatif merupakan penelitian non-hipotesis sehingga langkah penelitiannya tidak menggunakan rumusan hipotesis. Penggunaan pendekatan ini bertujuan mengembangkan tes yang baik.

Uji coba soal dilakukan dengan cara ujian (testing) setelah perangkat soal uji coba dikenakan pada responden. Data lembar jawab siswa merupakan data skunder yang bersifat dokumentatif. Lembar jawab siswa diperoleh langsung dari masing-masing SMP yang bersangkutan melalui uji coba.

Jumlah lembar jawab peserta tes yang digunakan untuk analisis butir sebanyak 800 lembar jawab untuk masing-masing kelas, seperti yang dikatakan oleh Depdiknas (2000: 7) bahwa untuk keakuratan hasil analisis butir soal, baik menurut teori klasik maupun modern, besarnya sampel peserta uji coba paling sedikit 750 peserta. Pengambilan lembar jawab sebanyak 800 ini juga telah memenuhi persyaratan yang dianjurkan oleh Nunnally (1978) bahwa jumlah responden untuk analisis butir minimal lima kali jumlah butir soal.

Rincian sekolah, peserta uji coba dalam penelitian ini dapat dilihat pada tabel berikut.

Tabel 1  
Jumlah Sekolah dengan Memperhatikan Mutu dan Lokasi Sekolah

Mutu Sekolah	Lokasi	
	Kodya/Kabupaten	Di luar Kodya/Kab
Baik	1	1
Sedang	2	2
Kurang	1	1
Jumlah	4	4

Sumber: Depdiknas, 2000

Masing-masing sekolah diberi lembar jawab sebanyak 100 untuk masing-masing perangkat tes, sehingga jumlah lembar jawab untuk setiap perangkat adalah 800.

## Teknik Analisis Butir Soal

### 1. Analisis Butir Soal secara kualitatif

Analisis kualitatif merupakan analisis data melalui penelaahan butir soal pada perangkat tes. Penelaahan secara kualitatif antara lain dilakukan untuk menyeleksi apakah suatu soal diperkirakan akan berfungsi dengan baik dan untuk mengetahui kehomogenan soal. Analisis kualitatif bertujuan untuk menilai butir soal secara teoritis yang dikaji dari sudut pandang isi atau materi tes, bahasa dan teknik penulisan soal (Depdiknas, 2000: 3).

Pada setiap perangkat, tes dianalisis oleh 3 orang penelaah. Penelaahan dilakukan dengan cara mencocokkan butir soal dengan kriteria-kriteria yang ada pada kartu telaah. Setiap penelaah cukup membubuhkan tanda *check* (P) ada kolom yang telah disediakan.

Tabel 2  
Kartu Telaah Butir Soal

Bidang Penelaahan	Kriteria Penelaahan	Ya	Tdk
Materi	<ol style="list-style-type: none"><li>1. Soal sesuai dengan indikator</li><li>2. Pengecoh sudah berfungsi</li><li>3. Hanya ada satu kunci jawaban</li></ol>		
Konstruksi	<ol style="list-style-type: none"><li>4. Pokok soal dirumuskan dengan singkat, jelas dan tegas</li><li>5. Rumusan pokok soal tidak mengarah memberi petunjuk ke kunci jawaban</li><li>6. Pokok soal bebas dari pernyataan yang bersifat negatif ganda.</li><li>7. Gambar/grafik/tabel/diagram dan sejenisnya jelas dan berfungsi</li><li>8. Pilihan jawaban homogen dan logis</li><li>9. Panjang pilihan jawaban relatif sama</li><li>10. Pilihan jawaban tidak menggunakan pernyataan "semua jawaban di atas benar" atau "semua jawaban di atas salah"</li><li>11. Pilihan jawaban yang berbentuk angka disusun berdasarkan urutan besar kecilnya angka</li><li>12. Butir soal tidak tergantung pada jawaban sebelumnya</li></ol>		
Bahasa	<ol style="list-style-type: none"><li>13. Soal menggunakan bahasa yang sesuai dengan kaidah Bahasa Indonesia</li><li>14. Soal menggunakan bahasa yang komunikatif</li><li>15. Soal tidak menggunakan bahasa yang berlaku setempat</li><li>16. Pilihan jawaban tidak mengulang kata/kelompok kata yang sama yang bukan merupakan satu kesatuan</li></ol>		

Sumber: Depdiknas, 2000

## 2. Analisis Butir Soal Secara Kuantitatif

Analisis kualitatif berfungsi untuk menguji apakah butir soal diperkirakan akan berfungsi dengan baik secara teoritis, sedangkan analisis kuantitatif untuk menguji apakah soal telah berfungsi dengan baik berdasarkan bukti empiris.

### a. Analisis butir dengan program *ITEMAN*

Analisis kuantitatif pada penelitian ini menggunakan program *ITEMAN* (*Item and Test Analysis*) versi 3.0 untuk pendekatan teori tes klasik (*Classical Test Theory/CTT*) dan program *BIGSTEPS* versi 2.30 untuk pendekatan teori respons butir (*Item Response Theory/IRT*).

Menurut teori tes klasik, secara empiris mutu butir soal ditentukan oleh statistik butir soal, yang meliputi tingkat kesukaran, daya beda, dan efektifitas distraktor. Berdasarkan statistik butir, kualitas butir soal dikategorikan baik, cukup baik dan tidak baik dengan ketentuan sebagai berikut: kategori baik, apabila tingkat kesukaran memiliki skor antara  $0,25 \leq p \leq 0,75$ , korelasi biserial ( $r_{bis}$ ) butir  $\geq 0,30$  dan korelasi biserial ( $r_{bis}$ ) pilihan jawaban bernilai negatif selain kunci.

### b. Analisis butir dengan program *BIGSTEPS*

Analisis butir dengan program *BIGSTEPS* ini dilakukan untuk menguji butir berdasarkan teori respons butir. Butir-butir tes yang dianalisis adalah seluruh butir soal yang ada pada perangkat tes tanpa memperhatikan hasil analisis sebelumnya. Dengan menggunakan program *BIGSTEPS* akan diketahui karakteristik butir yang berupa tingkat kesukaran butir. Parameter tingkat kesukaran butir yang terletak pada skala logits  $-2,00 \leq b_i \leq 2,00$  termasuk butir yang memiliki tingkat kesukaran sedang, butir soal dengan  $b_i < -2,00$  berarti sangat mudah, dan butir soal dengan nilai  $b_i > 2,00$  berarti sangat sukar.

Berdasarkan pendekatan teori respons butir model logistik satu parameter, kualitas butir soal dikelompokkan dalam kategori baik, cukup baik dan tidak baik dengan ketentuan sebagai berikut:

- 1) Baik, apabila butir cocok dengan model, dan parameter tingkat kesukaran  $-2,00 \leq b_i \leq 2,00$ .
- 2) Tidak baik, apabila butir tidak cocok dengan model.

### 1. Peserta Uji Coba

Depdiknas mensyaratkan untuk keperluan analisis empiris maka jumlah peserta yang dibutuhkan paling sedikit 750 siswa, yang secara merata terdiri dari siswa pandai, sedang dan siswa kurang pandai. Perbandingan antara sekolah pada katagori baik, sedang dan kurang adalah 25% : 50% : 25%. Dalam menentukan peserta uji coba peneliti juga memperhatikan letak geografis sekolah di dalam kota dan di luar kota kabupaten (Depdiknas, 2000).

Gambaran sekolah dan jumlah peserta yang dijadikan sebagai subjek uji coba soal baik untuk paket A maupun B dapat dilihat pada Tabel 1.

Tabel 3  
Nama Sekolah, Mutu, dan Jumlah Peserta Uji Coba

Mutu Sekolah	Nama Sekolah	Jumlah Peserta	
		Paket A	Paket B
Baik	SMPN 1 Wonosobo	120	111
	SMPN 2 Wonosobo	103	105
Sedang	SMPN 1 Mojotengah	115	117
	SMP Muhammadiyah 1 Wonosobo	80	83
	SMPN 1 Garung	57	54
	SMP Takhassus Al-Qur'an	123	124
Kurang	SMPN 2 Mojotengah	76	68
	SMPN 2 Garung	47	47
	SMP Islam Wonosobo	50	50
	Jumlah	771	759

Data lembar jawab peserta pada lampiran 3

## Hasil Penelitian dan Pembahasan

### Dengan Program Iteman

#### 1. Paket A

Melalui analisis dengan program Iteman terhadap soal paket A yang direspons oleh 771 siswa, diperoleh karakteristik butir untuk tingkat kesukaran, daya pembeda, dan keberfungsian distraktor sebagai berikut.

Tabel 4  
Jumlah Butir Baik, Cukup Baik, dan Tidak Baik  
Berdasarkan Keseluruhan Karakteristik

Butir baik		
No. butir	Jumlah butir	Persen
1, 2, 4, 7, 8, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 45, 48, 49	35	70%
Butir cukup baik		
No. butir	Jumlah butir	Persen
3, 5, 6, 9, 14, 18, 28, 29, 31, 41, 44, 46, 47, 50	14	28%
Butir tidak baik		
No. butir	Jumlah butir	Persen
40	1	2%

Berdasarkan hasil analisis, butir soal yang tidak baik sehingga perlu dibuang adalah butir nomor 40. Selanjutnya nomor butir ini tidak akan diikutsertakan pada analisis lanjutan dengan program Bigsteps.

## 2. Paket B

Melalui analisis dengan program Iteman terhadap soal paket B yang direspons oleh 759 siswa, diperoleh karakteristik butir untuk tingkat kesukaran, daya pembeda, dan keberfungsian distraktor disajikan pada Tabel 5:

Tabel 5  
Jumlah butir baik, cukup baik, dan tidak baik  
berdasarkan keseluruhan karakteristik

Butir baik		
No. butir	Jumlah butir	Persen
2, 4, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 23, 24, 25, 26, 27, 28, 29, 31, 32, 34, 35, 37, 38, 39, 40, 42, 44, 45, 47, 48, 49, 50	34	68%
Butir cukup baik		
No. butir	Jumlah butir	Persen
5, 7, 9, 10, 19, 20, 22, 30, 33, 36, 41, 43, 46	13	26%
Butir tidak baik		
No. butir	Jumlah butir	Persen
1, 3, 21	3	6%

Berdasarkan hasil analisis dan uji keputusan di atas maka butir soal yang tidak baik sehingga perlu dianulir adalah butir nomor 1, 3, dan 21. Selanjutnya nomor butir ini tidak akan diikutsertakan pada analisis lanjutan dengan program Bigsteps.

#### Dengan Program BIGSTEPS

##### 1. Soal Paket A

Berdasarkan analisis soal paket A diperoleh informasi bahwa dari 771 peserta yang tidak sesuai dengan model ada 18 peserta yaitu nomor peserta 84, 170, 374, 87, 37, 407, 58, 334, 76, 90, 134, 94, 405, 81, 394, 35, 150, dan 100. Butir soal yang tidak sesuai ada 3 dari 50 butir yang ada, yaitu butir nomor 40, 6 dan 18 (deleted Bigsteps). Butir soal yang cocok dengan model berarti butir soal tersebut berperilaku konsisten dengan apa yang diharapkan oleh model. Butir yang dianggap tidak sesuai dengan model adalah butir yang memiliki outfit  $> 2,00$ , atau daya pembedanya bernilai negatif.

Butir soal yang tidak cocok dengan model adalah butir nomor 6, 18, dan 40 (3 butir) dengan keterangan seperti pada Tabel 6. Tingkat kesukaran pada model Rasch merupakan parameter utama analisis kuantitatif. Tingkat kesukaran dinyatakan dengan skala logit (*log-odd-units*). Tingkat kesukaran ideal membentang dari  $-2,00$  sampai dengan  $2,00$ , sedang di bawah  $-2,00$  tergolong terlalu mudah, lebih dari  $2,00$  tergolong terlalu sukar.

Hasil analisis pada paket soal menunjukkan bahwa tingkat kesukaran soal membentang antara  $-3,05$  sampai dengan  $1,96$  logit dengan rincian sebagai berikut.

Tabel 6  
Butir Soal yang Tidak Cocok dengan Model

Nomor butir	<i>oufit</i>	<i>ptbis</i>	Keterangan	Kesimpulan
6	2,4	-0,02	<i>Outfit</i> $> 2,00$ ; <i>ptbis</i> negatif	TC
18	0,9	-0,01	<i>ptbis</i> negatif	TC
40	Dianulir (didrop) lewat analisis program IteMan			TC

Sumber: tabel 14.1. output2

Keterangan: TC = Tidak Cocok

Tingkat kesukaran tinggi : tidak ada

Tingkat kesukaran sedang : 43 butir

Tingkat kesukaran rendah : 4 butir

Besarnya daya pembeda dapat dilihat pada harga korelasi poin biserial yang idealnya bernilai positif. Dari hasil analisis, rentangan daya pembeda paket soal berkisar antara  $0,07$  sampai dengan  $0,48$  logit. Makin tinggi nilai daya pembeda maka makin baik soal tersebut dalam membedakan kelompok peserta tinggi (pandai) dengan kelompok peserta rendah (tidak pandai).

## 2. Soal Paket B

Berdasarkan analisis soal paket A diperoleh informasi bahwa dari 759 peserta, yang tidak sesuai dengan model ada 13 peserta yaitu peserta dengan nomor 169, 147, 97, 169, 315, 331, 115, 12, 166, 153, 37, 154, dan 34. Butir

soal yang tidak sesuai ada 4 dari 50 butir yang ada, yaitu butir nomor 1, 3, 21 (deleted pada program Iteman), dan 18 (deleted pada program Bigsteps).

Butir soal yang cocok dengan model berarti butir soal tersebut berperilaku konsisten dengan apa yang diharapkan oleh model. Butir yang dianggap tidak sesuai dengan model adalah butir yang memiliki outfit  $>2,00$  atau daya pembedanya bernilai negatif.

Butir soal yang tidak cocok dengan model adalah butir nomor 1, 3, 21, dan 34 (4 butir) dengan keterangan seperti pada Tabel 7.

Tabel 7  
Butir Soal yang Tidak Cocok dengan Model

Nomor butir	outfit	ptbis	Keterangan	Kesimpulan
34	2,4	0,7	outfit $>2,00$	TC
1	Dianulir (didrop) lewat analisis program Iteman			TC
3	Dianulir (didrop) lewat analisis program Iteman			TC
21	Dianulir (didrop) lewat analisis program Iteman			TC

Sumber: tabel 14.1. output2

**Keterangan:** TC = Tidak Cocok

Tingkat kesukaran pada model Rasch merupakan parameter utama analisis kuantitatif. Tingkat kesukaran dinyatakan dengan skala logit (*log-odds-unity*). Tingkat kesukaran ideal membentang dari  $-2,00$  sampai dengan  $2,00$ , sedang di bawah  $-2,00$  tergolong terlalu mudah, lebih dari  $2,00$  tergolong terlalu sukar.

Hasil analisis pada paket soal menunjukkan bahwa tingkat kesukaran soal membentang antara  $-2,27$  sampai dengan  $3,30$  logit dengan rincian keterangan:

Tingkat kesukaran tinggi : 2 butir

Tingkat kesukaran sedang : 43 butir

Tingkat kesukaran rendah : 1 butir

Besarnya daya pembeda dapat dilihat pada harga korelasi point biserial yang idealnya bernilai positif. Berdasarkan hasil analisis, rentangan daya pembeda

paket soal berkisar antara 0,08 sampai dengan 0,45 logit. Semakin tinggi harga daya pembeda maka semakin baik soal tersebut.

### Fungsi Informasi

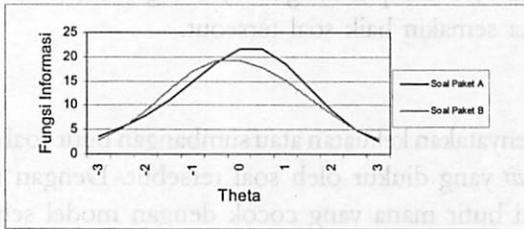
Fungsi informasi menyatakan kekuatan atau sumbangan butir soal dalam mengungkapkan *latent trait* yang diukur oleh soal tersebut. Dengan fungsi informasi dapat diketahui butir mana yang cocok dengan model sehingga membantu dalam seleksi butir soal. Fungsi informasi soal Paket A dan Paket B beserta harga *standard error of measurement (SEM)* pada rentangan kemampuan  $-3,0 \leq \theta \leq 3,0$  dengan interval 0,5 dapat dilihat pada Lampiran 7 yang diringkas pada Tabel 8.

Pada Tabel 8 dapat dilihat bahwa nilai fungsi informasi tertinggi pada  $q=0,5$ , untuk soal Paket A dan  $q=0,0$  untuk soal Paket B. Perbandingan lengkungan grafik untuk fungsi informasi soal Paket A dan Paket B dapat dilihat pada Gambar 1.

Tabel 8  
Harga Fungsi Informasi dan SEM

$\theta$	Paket A		Paket B	
	I( $\theta$ )	SEM	I( $\theta$ )	SEM
-3,0	3,486	0,536	2,849	0,592
-2,5	5,463	0,428	5,397	0,430
-2,0	7,954	0,355	9,014	0,300
-1,5	11,070	0,301	13,186	0,275
-1,0	14,892	0,259	16,956	0,243
-0,5	18,820	0,231	19,154	0,228
0,0	21,582	0,215	19,265	0,228
0,5	21,657	0,215	17,729	0,237
1,0	18,525	0,232	15,048	0,258
1,5	13,693	0,270	11,64	0,293
2,0	8,918	0,335	8,074	0,352
2,5	5,090	0,443	5,151	0,441
3,0	2,572	0,623	3,164	0,562

Sumber: lampiran 6



Gambar 1. Lengkungan Fungsi Informasi Soal Paket A dan B

### Relatif Efisiensi

Relatif efisiensi merupakan perbandingan fungsi informasi soal Paket A dengan Paket B. Relatif efisiensi tergantung pada nilai  $q$ . Pada wilayah tertentu mungkin saja soal Paket A lebih efektif terhadap Paket B, tetapi dapat juga sebaliknya.

Pada rentangan  $-3,0 \leq q \leq 3,0$  dengan interval 0,5, nilai relatif efisiensi soal Paket A terhadap Paket B dapat dilihat pada Tabel 9.

Tabel 9  
Nilai Relatif Efisiensi Soal Paket A Terhadap Paket B

Skala Kemampuan	Relatif Efisiensi (RE)
-3,0	1,224
-2,5	1,012
-2,0	0,882
-1,5	0,840
-1,0	0,878
-0,5	0,983
0,0	1,120
0,5	1,222
1,0	1,231
1,5	1,178
2,0	1,105
2,5	0,988
3,0	0,813

Pada semua harga  $q$  ( $-3,0 < q < 3,0$ ), relatif efisiensi soal Paket A terhadap Paket B adalah sebesar:

$$RE = \frac{I_A}{I_B} = \frac{757,93}{722,43} = 1,049$$

Berdasarkan uraian di atas, selanjutnya dapat dikatakan bahwa soal Paket A lebih efektif 1,049 kali soal Paket B. Agar relatif efisiensinya sama maka soal Paket B panjangnya harus 1,049 kali soal paket A, atau bila paket A 50 butir soal maka paket B harus 52 butir soal.

### Penyetaraan Paket Soal

Penyetaraan paket soal pada teori respons butir ini menggunakan metode rerata dan sigma. Metode ini menguntungkan karena dalam penentuan konstanta konversi bersifat timbal balik, berbeda dengan metode regresi yang bersifat searah (Naga, 1992: 375).

Menurut Hambleton dan Swaminathan (1985), hubungan antara estimasi parameter butir tes atau estimasi kemampuan peserta pada kedua perangkat tes yang akan disetarakan memenuhi hubungan:

$$y = \alpha x + \beta$$

$$\bar{y} = \alpha \bar{x} + \beta$$

$$\alpha = \frac{S_y}{S_x}$$

$$\beta = \bar{y} - \alpha \bar{x}$$

#### Keterangan:

$y$  = estimasi kemampuan atau estimasi parameter butir pada tes Y

$x$  = estimasi kemampuan atau estimasi parameter butir pada tes X

$\bar{y}, \bar{x}$  = rerata dari  $y$  dan  $x$

$S_y, S_x$  = standar deviasi dari  $x$  dan  $y$

Proses penyetaraan dilakukan dengan menggunakan parameter tingkat kesukaran pada *anchor item* yang jumlahnya 5 butir. Harga parameter tingkat kesukaran pada masing-masing paket dapat dilihat pada Tabel 10.

Tabel 10  
Parameter Tingkat Kesukaran Pada *Anchor Item*

No. butir	Paket A	Paket B
17	-0,25	-0,16
24	0,58	0,09
27	-0,39	-0,32
37	0,22	-0,38
41	-3,05	-2,27
Rerata	-0,578	-0,608
Simpangan baku	1,283081	0,846768

Tabel 10 menunjukkan bahwa:

$$\bar{x} = -0,578 \quad S_x = 1,283081$$

$$\bar{y} = -0,608 \quad S_y = 0,846768$$

$$a = \frac{0,84676}{1,28308} = 0,659949$$

$$b = -0,608 - [0,659949(-0,578)] = -0,226549$$

Kemudian diperoleh formula konversi:

$$y = 0,660 x - 0,227$$

$$b_B = 0,660 b_A - 0,227$$

**Keterangan:**  $b_A, b_B$  = tingkat kesukaran soal paket A dan B

Berdasarkan formula konversi di atas, tingkat kesukaran soal paket A lebih tinggi daripada soal paket B.

## **Simpulan**

Setelah melalui tahapan-tahapan penelitian dan pembahasan, maka pada akhir bab ini dapat dibuat kesimpulan sebagai berikut:

Hasil analisis empiris menunjukkan bahwa:

1. Rerata tingkat kesukaran soal paket A adalah 0,45 logit, sedang soal paket B adalah 0,32 logit. Ini berarti soal paket A lebih sukar dibandingkan dengan soal paket B.
2. Besarnya daya pembeda untuk soal paket A antara 0,07 – 0,48 logit, sedangkan untuk soal paket B antara 0,08 – 0,45 logit. Dengan demikian butir-butir soal pada kedua paket tersebut dapat diterima, karena tidak bernilai negatif.
3. Indeks reliabilitas berdasarkan teori tes klasik untuk soal paket A dan paket B sama yaitu sebesar 0,99, berarti keduanya tergolong sangat baik.
4. Fungsi informasi maksimum untuk soal paket A adalah  $q=0,0$  dan untuk soal paket B adalah  $q=0,5$ . Hal ini berarti kedua paket soal sangat cocok digunakan pada siswa yang memiliki tingkat kemampuan menengah atau sedang.
5. Pada semua harga  $q(-3,0 \leq q \leq 3,0)$ , nilai relatif efisiensi soal Paket A terhadap Paket B adalah sebesar 1,049, sehingga dapat dikatakan bahwa soal Paket A lebih efektif 1,049 kali soal Paket B.
6. Formula konversi sekor untuk penyetaraan soal paket A ke dalam soal paket B adalah  $B=0,660 A - 0,227$ . Ini berarti soal paket A lebih sukar daripada soal paket B.
7. Dengan melihat karakteristiknya, secara umum paket soal yang disusun dapat disimpan ke bank soal dalam rangka pengembangan bank soal di Kabupaten Wonosobo.
8. Butir soal paket A yang relevan dimasukkan ke bank soal sebanyak 47 butir dari 50 butir (94%) yang diujicobakan, sedangkan butir soal paket B yang relevan dimasukkan ke bank soal sebanyak 46 butir dari 50 butir (92%) yang diujicobakan.
9. Nomor butir yang tidak relevan untuk soal paket A adalah 6, 18, dan 40, sedangkan untuk soal paket B adalah 1, 3, 21, dan 34. Butir-butir ini tidak

relevan karena statistik outfit tidak cocok dengan model, yaitu  $>2,00$  logit dengan harga pbtis yang menyatakan daya pembeda bernilai negatif.

Soal Paket A dan Paket B keduanya dapat diangkat ke dalam bank soal di Kabupaten Wonosobo. Diperoleh 94% butir soal Paket A dan 92% butir untuk paket B yang sesuai dengan persyaratan bank soal.

### **Saran-saran**

Berdasarkan simpulan penelitian di atas disarankan hal-hal sebagai berikut.

1. Prosedur pengembangan bank soal pada penelitian ini disarankan dikembangkan untuk mata pelajaran yang lain.
2. Prosedur kalibrasi butir tes dapat dikembangkan dengan menggunakan model 2 parameter.
3. Butir-butir soal yang tidak fit sebaiknya diganti agar validitas isi tes tetap terjamin.
4. Perlu dilakukan penelitian tentang perbandingan pengembangan bank soal dengan teori tes klasik dan dengan teori tes modern.

### **Daftar Pustaka**

- Allen, J.M., & Yen, M.W. (1979). *Introduction to measurement theory*. Monterey: Brook/Cole Publishing Company.
- Depdiknas. (2000). *Penilaian dan pengujian untuk guru SLTP*. Jakarta: Dinas Pendidikan Nasional.
- \_\_\_\_\_. (2000). *Pedoman pengembangan bank soal*. Jakarta: Departemen Pendidikan Nasional.
- Djemari Mardapi. (1998). Pengembangan tes intelegensi untuk tes masuk ke perguruan tinggi. *Jurnal Kependidikan*. 25-34.
- Hambleton, R.K., Swaminathan, H., & Rogers H.J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications, Inc.
- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. London: Sage Publications Inc.
- Muhamad Noer. (1987). *Pengantar teori tes*. Jakarta: Dirjen Dikti Depdikbud.