

## **KARAKTERISTIK BUTIR SOAL ULANGAN KENAIKAN KELAS SEBAGAI PERSIAPAN *BANK* SOAL BAHASA INGGRIS**

*Nur Hidayanto Pancoro SP*  
Pendidikan Bahasa Inggris FBS UNY  
Jl. Colombo No.1 Yogyakarta 55281  
noersabar@yahoo.com

### **Abstrak**

Penelitian ini bertujuan untuk mengembangkan rintisan bank soal Bahasa Inggris SMP kelas VIII pada tingkat provinsi dengan memperhatikan (1) karakteristik soal Ulangan Kenaikan Kelas (UKK); (2) karakteristik butir-butir soal; (3) jumlah butir soal yang memenuhi standar. Metode yang digunakan adalah pengembangan dengan pendekatan kuantitatif deskriptif. Sampel sebanyak 6554 respon siswa (lembar jawab). Perangkat tes dianalisis kesesuaian indikatornya dengan Standar Kompetensi (SK) dan Kompetensi Dasar (KD) yang telah ditentukan, serta ditelaah secara kualitatif oleh ahli berdasarkan kriteria konstruksi, bahasa dan materi. Respons siswa dianalisis secara kuantitatif dengan *Iteman 3.00*, dan *Bilog MG 3.0 3 PL* untuk mendapatkan informasi tentang tingkat kesulitan, daya beda, *pseudoguessing*, serta informasi butir. Bank soal yang dihasilkan masuk kategori baik dilihat dari (1) Hasil analisis kualitatif dari 150 butir soal, sebanyak 105 butir soal (70%) dinyatakan baik. (2) Hasil analisis kuantitatif dengan program *Iteman 3.00* diperoleh 115 butir soal tergolong baik. (3) Analisis dengan *Bilog MG 3 PL* 77 butir soal (66,96%) masuk kategori baik.

Kata kunci: *bank soal, tingkat kesulitan, daya beda, pseudo guessing, expert judgment, informasi butir dan tes*

## **THE ITEM CHARACTERISTICS OF THE FINAL SEMESTER TEST AS A PREPARATION FOR ENGLISH ITEM BANK**

*Nur Hidayanto Pancoro SP*  
English Education Department FBS UNY  
Jl. Colombo No.1 Yogyakarta 55281  
noersabar@yahoo.com

### **Abstract**

This research aims at developing a pilot english test item bank of the 2<sup>nd</sup> grade junior high school based on: (1) the characteristics of test items for final semester test; (2) the total number of test items for the 2<sup>nd</sup> grade junior high school in Yogyakarta Special Province. This descriptive quantitative research used 6554 samples of students' answer sheets in 2009 and also employs expert judgment to review the construct, language, as well as the materials of the questions. Students' responses were also analyzed by using *Iteman 3.00*, *Bilog MG 3.0 3 PL*. The result of qualitative analysis with expert judgment which shows that 105 out of the 150 test items are categorized as "good", and the other are "bad". The quantitative analysis with *Iteman 3.00* shows that 115 out of the 150 test items are categorized as "good" and the quantitative analysis with *Bilog MG 3.00 3 PL* shows that 67% test items are categorized as good.

Key words: *item bank, level of difficulty, discrimination index, pseudoguessing, expert judgment, item and test information function*

## **Pendahuluan**

Bahasa Inggris, sebagai salah satu mata pelajaran wajib yang di UN-kan, bertujuan untuk membekali peserta didik dengan kemampuan untuk berkomunikasi dan memahami serta mengungkapkan makna baik dalam percakapan transaksional dan interpersonal dalam Bahasa Inggris (KTSP, 2006). Tujuan tersebut merupakan salah satu dari seluruh tujuan yang diamanatkan dalam KTSP 2006 untuk SMP. Kemampuan-kemampuan berkomunikasi tersebut kemudian dirinci ke dalam standar kompetensi yang idealnya dicapai setelah proses pembelajaran. Pencapaian hasil belajar dicerminkan dalam bentuk indikator-indikator pembelajaran yang selanjutnya dapat dievaluasi untuk mengetahui tingkat keberhasilan pembelajaran.

Keberhasilan pembelajaran Bahasa Inggris tidak bisa terlepas dari adanya evaluasi yang tepat (Brown, 2001: 2). Ulangan Kenaikan Kelas (UKK) merupakan salah satu barometer pengukuran keberhasilan pembelajaran Bahasa Inggris selama satu semester. Kualitas soal yang digunakan untuk mengetes siswa sangat mempengaruhi kualitas hasil pengukuran yang pada akhirnya juga mempengaruhi hasil evaluasi pendidikan. Dengan kata lain, untuk mendapatkan hasil evaluasi pembelajaran yang baik, tes yang digunakan dalam UKK pun juga harus terjamin kualitasnya.

Dalam pelaksanaannya, tidak semua sekolah mampu untuk menyusun Ulangan Kenaikan Kelas bagi sekolah mereka sendiri. Untuk itu, MGMP kabupaten di masing-masing kabupaten ataupun kotamadya diberikan hak untuk menyusun perangkat UKK. Di DIY sendiri, penyusunan soal UKK telah diserahkan kepada masing-masing MGMP kabupaten, yang terdapat di 4 kabupaten dan 1 kotamadya. Meskipun soal UKK yang disusun oleh MGMP kabupaten/kotamadya bersumber dari standar isi yang sama, karakteristik soal yang dihasilkan berbeda. Untuk itu, soal-soal tersebut perlu terlebih dahulu dianalisis agar memiliki karakteristik yang sama sehingga benar-benar dapat digunakan untuk mengukur mutu pendidikan. Dalam kenyataannya, belum dievaluasi juga apakah tes tersebut disusun atas butir-butir yang baik. Ketiadaan bank soal Bahasa Inggris yang baik untuk SMP juga merupakan salah satu faktor yang melatarbelakangi kesulitan dalam menyusun tes yang memiliki butir-butir yang baik. Dengan

keberadaan bank soal tersebut nantinya, standarisasi mutu pendidikan akan lebih mudah dilakukan dan dipantau.

Berdasarkan pemikiran di atas, keberadaan bank soal Bahasa Inggris merupakan suatu hal yang sangat diperlukan menimbang sebagian besar soal atau kalau boleh dikatakan seluruh soal yang telah digunakan untuk mengevaluasi pembedajaran Bahasa Inggris, khususnya wilayah provinsi, belum diketahui karakteristiknya. Untuk itu, rintisan bank soal yang disusun dengan menggunakan cara yang terstruktur dan terstandar sehingga menghasilkan butir-butir soal yang diketahui karakteristiknya, terkalibrasi dan tersetarakan, mutlak diperlukan untuk menjamin mutu pendidikan di wilayah Provinsi DIY. Berdasarkan permasalahan yang ada, maka tujuan penelitian ini antara lain sebagai berikut.

1. Mengetahui karakteristik butir soal UKK Bahasa Inggris dari Kabupaten Gunungkidul, Sleman, dan Kulonprogo untuk menentukan jumlah soal yang layak dimasukkan dalam rintisan bank soal Bahasa Inggris kelas VIII SMP.
2. Menyusun rintisan bank soal Bahasa Inggris SMP kelas VIII yang terkalibrasi dengan baik.

Penilaian hasil pembelajaran Bahasa Inggris sangat erat kaitannya dengan topik bahasan hingga kurikulum yang berlaku. Sebuah tes hasil belajar haruslah benar-benar mengukur apa yang seharusnya diukur, yakni materi atau bahan yang diberikan dalam kegiatan pembelajaran. Tes tersebut dapat diberikan dalam bentuk pilihan ganda, menjodohkan, melengkapi kalimat, hingga menjawab pertanyaan *essay*.

Bachman & Palmer (1996: 176) menambahkan bahwa sebuah perangkat tes Bahasa Inggris haruslah disusun untuk benar-benar mengukur pengetahuan dan 4 keterampilan berbahasa yakni *Speaking, Listening, Reading, dan Writing*. Untuk itu, penyusun tes perlu memutuskan kemampuan apa yang perlu diikutsertakan untuk dites, serta kemampuan mana yang tidak boleh diikutsertakan. Selama penyusunan tes, penyusun haruslah menspesifikasikan *theoretical definition* atau definisi teoritis dari konstruk yang ingin diukur. Dalam pelaksanaannya, definisi konstruk inilah

yang akan mencakup komponen khusus dari pengetahuan dan kemampuan berbahasa.

Sebagai salah satu langkah untuk mendukung keberhasilan proses evaluasi pembelajaran, keberadaan bank soal merupakan suatu hal yang sangat diperlukan. Menurut Depdikbud (1995: 1) bank soal merupakan suatu kumpulan butir-butir soal yang telah diketahui karakteristiknya dan telah dinyatakan valid serta reliabel. Dengan demikian, bank soal diperlukan untuk menunjang kegiatan penilaian karena dengan adanya bank soal, berarti guru dapat menggunakan butir-butir soal yang terdapat dalam bank soal tersebut. Butir soal yang telah diketahui karakteristiknya dan telah dinyatakan sesuai standar (valid dan reliabel) sangat membantu guru dalam hal mengukur dan menilai hasil pembelajaran.

Bank soal bukanlah sekedar kumpulan butir soal atau pertanyaan, namun merupakan butir-butir soal yang telah terkalibrasi sehingga dapat digunakan untuk menyediakan informasi mengenai *trait* peserta tes (Chang, 2005: 2). Dari definisi tersebut, butir-butir bank soal bukanlah hanya sekedar butir soal yang dibuat kemudian dimasukkan ke dalam buku kumpulan soal, namun haruslah diketahui karakteristik butir soalnya terlebih dahulu, dan dipastikan bahwa butir-butir soal tersebut andal.

Salah satu tujuan penyusunan sebuah bank soal adalah untuk keperluan pengembangan beberapa tes yang setara (Eid, 2005: 170). Alasannya adalah karena sebuah bank soal tersusun atas butir-butir soal yang sudah diketahui karakteristik yang dapat digunakan untuk menyusun tes atau subtes baru yang karakteristiknya dapat diprediksi.

Bank soal yang baik terdiri dari butir-butir soal yang telah dikalibrasikan. Butir-butir soal tersebut memberikan gambaran mengenai seberapa sulit butir soal tersebut, seberapa jauh butir-butir tersebut mampu membedakan siswa yang memiliki kemampuan lebih dengan siswa yang kemampuannya kurang, serta seberapa besar kemungkinan atau probabilitas setiap siswa dengan kemampuan apapun menjawab benar dengan menebak.

Seperti telah diungkap sebelumnya, butir-butir tes hasil UKK harus diseleksi terlebih dahulu sebelum dimasukkan ke dalam rintisan bank soal.

Butir soal yang dimasukkan ke dalam rintisan bank soal adalah butir soal yang memiliki karakteristik baik, yakni memenuhi beberapa syarat soal yang baik antara lain tingkat kesulitan, daya beda, dan proporsi pemilih.

Butir-butir soal yang baik paling tidak haruslah memenuhi 3 syarat, yakni tingkat kesulitan soal, daya beda, dan tiap pilihan minimal dipilih oleh 5% pemilih. Menurut Mardapi (2008: 143), butir soal yang baik memiliki kisaran indeks kesulitan 0,3 – 0,7. Butir soal yang memiliki tingkat kesulitan di bawah 0,3 dianggap terlalu sulit dan butir soal yang memiliki tingkat kesulitan di atas 0,7 dianggap terlalu mudah. Harga indeks daya beda butir soal yang boleh digunakan adalah  $\geq 0,3$ , dan tiap butir minimal dipilih 5% peserta tes.

Fernandes (1984: 25-29) menyatakan bahwa butir soal yang baik memiliki tingkat kesulitan antara 0,25-0,75, daya beda  $> 0,2$ . Syarat yang lain dari sebuah butir soal yang baik adalah proporsi pemilih yang kemudian dijadikan syarat distraktor yang baik. Sebuah butir soal yang baik setiap butir distraktornya harus dipilih minimal 2% peserta tes. Dalam penelitian ini, pedoman yang diberikan oleh Fernandes inilah yang digunakan untuk memutuskan butir soal yang baik.

Phase 3 program *BILOG MG* menunjukkan informasi mengenai daya beda butir (*item discriminating/slope*), tingkat kesulitan butir soal (*difficulty of item/threshold*), dan peluang tebakan semu (*pseudoguessing/lower asymptote estimate*). Nilai daya beda berkisar antara 0,0 – 2,0, tingkat kesulitan berkisar antara -2,0 – 2,0, dan *pseudoguessing* berkisar antara  $< 0,25$ . Nilai daya beda yang tinggi menunjukkan butir memiliki daya beda yang semakin bagus sedangkan nilai daya beda yang semakin rendah menunjukkan sebaliknya. Nilai tingkat kesulitan yang tinggi menunjukkan butir tersebut semakin sulit sedangkan semakin rendah nilai tingkat kesulitan, semakin mudah butir tersebut. Nilai *pseudoguessing* yang tinggi menunjukkan semakin besar kemungkinan menebak dengan jawaban benar, sedangkan semakin rendah nilai *pseudoguessing*, semakin kecil kemungkinan menebak jawaban dengan benar.

## **Metode Penelitian**

Penelitian ini mengambil tempat di tiga kabupaten di Provinsi DIY, yakni Gunungkidul, Sleman, dan Kulonprogo. Penelitian ini dilaksanakan selama 8 bulan, dimulai pada bulan Juli 2009 dan selesai pada bulan Februari 2010.

Penelitian ini merupakan penelitian yang menggunakan pendekatan kuantitatif deskriptif dengan jenis penelitian pengembangan untuk menyusun rintisan bank soal Bahasa Inggris SMP dengan basis Provinsi DIY.

Asumsi yang digunakan dalam penelitian ini adalah bahwa pelaksanaan UKK Bahasa Inggris SMP Kelas VIII di DIY terlaksana dengan baik dan bersih serta sesuai standar yang telah ditetapkan. Desain yang digunakan dalam penelitian ini adalah *concurrent equating design* atau lebih sering dikenal dengan penyetaraan bersama.

Prosedur yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan Perangkat Soal dan Kisi-kisi UKK Bahasa Inggris dari Tiga Kabupaten di DIY

Pengumpulan Perangkat Soal UKK Bahasa Inggris dari tiga kabupaten di DIY dilakukan oleh peneliti dengan cara meminta arsip soal UKK yang sudah ada dan digunakan sebelumnya dari sekolah atau instansi terkait. Perangkat soal tersebut merupakan perangkat soal yang disusun oleh MGMP tiga kabupaten di DIY.

2. Penyesuaian Butir Soal dalam Perangkat Soal dengan Standar Isi

Perangkat soal yang telah terkumpul kemudian dilihat butir-butirnya untuk melihat apakah semua butir soal sesuai dengan standar isi. Dengan demikian, butir soal yang diambil diharapkan benar-benar mengukur kompetensi siswa seperti yang dicantumkan dalam standar isi. Dengan kata lain, siswa yang dikatakan lulus ujian semester dapat dipastikan memiliki kompetensi dasar minimal level tersebut.

3. Validasi Butir Soal oleh Ahli (*Expert Judgment*)

Expert yang digunakan dalam penelitian ini adalah dosen-dosen Jurusan Pendidikan Bahasa Inggris yang sudah berkecimpung dan berpengalaman dengan standar isi serta evaluasi pendidikan. Pedoman pe-

nelaahan berupa format *check list* yang akan disediakan peneliti dengan acuan dari Pedoman Penulisan Soal dari Depdiknas. *Check list* digunakan untuk menelaah butir soal ditinjau dari aspek materi, konstruksi, dan bahasa.

#### 4. Kalibrasi Butir Soal Secara Empiris

Selain dianalisis atau ditelaah berdasarkan SK, KD, dan indikator serta telaah *expert judgement*, seluruh butir soal juga dianalisis dengan menggunakan dua program analisis data yang telah disebutkan diawal, yakni *Iteman 3.0* dan *Bilog MG*. Kedua program tersebut digunakan untuk mencari tingkat keandalan butir soal, apakah sudah memenuhi standar/sahih atau belum. Program *Bilog MG* digunakan juga untuk pengkalibrasian butir soal yang telah disusun.

#### 5. Penyusunan Rintisan Bank Soal

Butir-butir soal yang sudah dianalisis karakteristiknya dan telah disetarakan kemudian disusun kembali menjadi rintisan bank soal.

Populasi dalam penelitian adalah seluruh perangkat soal UKK Bahasa Inggris di 4 kabupaten dan 1 kotamadya di DIY. Sampel dalam penelitian ini adalah perangkat soal dan lembar jawab UKK Bahasa Inggris SMP kelas VIII di Kabupaten Gunungkidul, Sleman, dan Kulonprogo semester 2 tahun pelajaran 2008/2009, yang dalam hal ini peneliti mengambil 6554 sampel. Pertimbangan pengambilan sampel ini adalah untuk mencakup sebagian besar Kompetensi Dasar (KD) SMP yang telah ditetapkan. Dalam penelitian ini sampel data awal adalah lembar jawab UKK Bahasa Inggris siswa SMP di DIY yang telah diambil dengan menggunakan 3 kriteria yakni sekolah dengan prestasi UN atas, tengah, dan bawah yang dilihat dari data Puspendik.

### **Hasil Penelitian**

Hasil dari pengumpulan data berupa perangkat tes UKK Bahasa Inggris SMP kelas VIII tahun 2009 dan juga respon siswa dari Kabupaten Gunungkidul, Sleman, dan Kulonprogo yang tertuang dalam lembar jawab. Data yang berupa respon siswa dari Kabupaten Sleman sudah berupa *soft*

*file* hasil *scan*, sedangkan data respon siswa dari Kabupaten Gunungkidul, dan Kulonprogo masih berupa lembar jawab sehingga peneliti men-*scan* sendiri data tersebut untuk mendapatkan data berbentuk *soft file*. Data *soft file* berisi rekap opsi jawaban siswa terhadap pertanyaan pilihan ganda dalam format ABCD. Rekap data tersebut kemudian disusun dengan cara dijadikan satu file dengan bantuan *Editplus 2*. Penyusunan data tersebut dimulai dari data Kabupaten Gunungkidul, Sleman, dan terakhir Kabupaten Kulonprogo.

Setelah respon siswa terhadap pertanyaan pilihan ganda dikumpulkan dalam satu file dengan *Editplus 2*, langkah berikutnya adalah menelaah seluruh butir soal pilihan gandanya berdasarkan standar kompetensi, kompetensi dasar, dan indikatornya. Hal ini dilakukan untuk mengetahui konstruksi atau kisi-kisi soal yang dikembangkan. Sebelum ditelaah konstruksinya, seluruh butir soal pilihan ganda tersebut diurutkan yakni butir soal 1 sampai 50 merupakan butir soal pilihan ganda yang berasal dari Kabupaten Gunungkidul, butir soal 51 sampai 100 merupakan butir soal dari Kabupaten Sleman, dan butir soal 101 sampai 150 merupakan butir soal dari Kabupaten Kulonprogo.

Dari ke-150 butir soal tersebut 36 butir (24%) merupakan pertanyaan informasi rinci dari teks, 30 butir (20%) merupakan pertanyaan yang berupa makna kata, frasa ataupun kalimat, 25 butir (16,67%) merupakan pertanyaan yang berupa isian titik-titik berupa kata ataupun frasa berdasarkan konteks kalimat dalam paragraf, 20 butir (13,33%) merupakan pertanyaan yang berupa kata ataupun kalimat, 15 butir pertanyaan (10%) merupakan pertanyaan informasi faktual, 14 butir pertanyaan (9,33%) berupa tujuan komunikatif teks, 6 butir pertanyaan (4%) berupa pertanyaan tentang gagasan utama, serta 4 butir pertanyaan (2,67%) pertanyaan berupa pertanyaan tentang informasi pendukung.

Hasil telaah oleh ahli (*expert*) dirangkum dan ditentukan tingkat kualitas butir tes yang memenuhi kriteria sebagai berikut:

1. Butir soal diterima yaitu butir soal yang memenuhi semua kriteria yang ada.

2. Butir soal yang ditolak yaitu butir soal yang tidak memenuhi salah satu kriteria pada ketiga aspek tersebut.

Untuk melakukan telaah terhadap seluruh butir soal tersebut, peneliti telah meminta bantuan 3 dosen dari Jurusan Pendidikan Bahasa Inggris UNY, yakni Drs. Suharso, M.Pd., Dra. Jamilah, M.Pd., dan Ari Purnawan, M.Pd., M.A. Hasil *expert judgment* menunjukkan bahwa dari 150 soal UKK Bahasa Inggris kelas VIII di 3 kabupaten di DIY, 106 butir soal (70,66%) dinyatakan baik oleh ke 3 ahli sedangkan 44 butir lainnya (29,3%) dinyatakan kurang baik. Dengan demikian, 106 butir soal tersebut telah memenuhi kaidah penyusunan butir soal pilihan ganda dalam hal materi, konstruksi dan bahasa.

Ke-44 butir soal yang dinyatakan kurang baik oleh ahli tersebut merupakan butir-butir yang tidak memenuhi kaidah materi, bahasa dan konstruksi. Dari ke-44 butir tersebut, 17 butir soal (38,64%) kurang memenuhi kaidah penyusunan butir soal pilihan ganda dalam hal panjang pilihan jawaban atau opsi jawaban yang tidak sama.

Jumlah butir soal yang pilihan jawabannya tidak homogen dan logis adalah 10 butir atau 22,72%. Butir soal tersebut antara lain 8, 31, 32, 33, 34, 35, 37, 50, 58, dan 113. Jumlah butir soal yang tidak menggunakan kaidah bahasa Inggris yang baik adalah 9 butir (20,45%). Butir-butir soal tersebut antara lain butir soal nomor 20, 120, 121, 123, 129, 132, 144, 148, dan 118. Butir-butir soal tersebut tidak menggunakan kaidah bahasa Inggris dalam pertanyaan ataupun pilihan jawabannya.

Jumlah butir soal yang pilihan jawabannya berbentuk angka tidak diurutkan besar kecilnya angka tersebut adalah 3 butir (6,82%). Butir-butir soal tersebut antara lain 40, 72, dan 112. Jumlah butir soal yang pokok soalnya tidak dirumuskan dengan jelas, singkat, dan tegas adalah 2 butir (4,44%). Butir-butir soal tersebut antara lain soal nomor 12 dan 99. Jumlah butir soal yang kunci jawabannya lebih dari 1 adalah 2 buah butir soal (4,55%) yakni butir soal nomor 54 dan 89. Jumlah butir soal yang menggunakan gambar dan sejenisnya namun tidak jelas dan berfungsi dengan baik adalah 1 butir soal (2,27%) yakni butir soal nomor 111 yang merupakan butir soal ke 11 dari Kabupaten Kulonprogo.

Dalam penelitian ini, karakteristik perangkat tes UKK Bahasa Inggris kelas VIII SMP TA 2008/2009 dicari dengan menggunakan bantuan program komputer *IteMan 3.00* untuk menganalisis dengan teori tes klasik serta *BILOG MG 3.0* untuk mengetahui karakteristik perangkat tes dengan menggunakan teori respon butir. Dari hasil analisis dengan *IteMan 3.00* terhadap 106 butir soal UKK Bahasa Inggris kelas VIII yang berdasarkan *expert judgment* dikategorikan baik, diperoleh informasi bahwa 67 butir soal baik (63,21%), sedangkan 39 butir soal kurang baik (36,79%), dengan perincian 14 butir soal (13,21%) memiliki tingkat kesulitan sekaligus daya beda yang tidak memenuhi syarat butir soal yang baik, 17 butir soal (16,04%) butir soal yang tidak mampu membedakan siswa kemampuan tinggi dan rendah, serta 8 butir soal (7,55%) butir soal yang tingkat kesulitannya tidak memenuhi syarat butir soal yang baik.

Analisis perangkat tes dengan *IteMan 3.00* juga menghasilkan informasi mengenai perangkat tes secara keseluruhan. Berdasarkan hasil analisis empiris diperoleh bahwa indeks reliabilitas (nilai alpha) untuk tes ini sebesar 0,723. Tes dengan indeks reliabilitas minimal 0,70 termasuk tes yang sudah reliabel. Nilai indeks reliabilitas berkaitan erat dengan kesalahan pengukuran (*Standard Error Measurement, SEM*). Besaran ini menunjukkan tingkat ketelitian hasil pengukuran. Pada tes ini diperoleh nilai SEM sebesar 4,702.

Hasil analisis empiris menunjukkan bahwa rerata skor tampak peserta adalah 48,912 dengan nilai simpangan baku sebesar 8,937 dan rentang skor dari 26,00 sampai dengan 86,00. Distribusi skor peserta membentuk kurva yang sedikit juling ke kanan. Hal ini dapat dilihat dari harga rerata yang lebih besar dari median ( $48,912 < 48,000$ ) dan harga *skew* sebesar -0,314. Kurva yang juling ke kanan menunjukkan bahwa sebagian besar peserta mendapat skor yang tinggi.

Seperti disampaikan pada bab sebelumnya, untuk mengetahui karakteristik butir soal dengan teori respon butir digunakan bantuan program komputer *Bilog MG*. Sebelum dilakukan analisis dengan program ini, peneliti terlebih dahulu menentukan model mana yang lebih tepat digunakan untuk menganalisis data dengan menggunakan metode plot kurva karakteristik butir. Metode ini dilakukan dengan melihat plot kurva

ICC yang dihasilkan dalam *output* hasil analisis Bilog MG dengan 1, 2, dan 3 parameter. Plot tersebut menunjukkan seberapa banyak *outlier*-nya. Butir yang dianggap cocok adalah butir yang probabilitasnya lebih besar atau sama dengan 0,01 atau 1%. Hasil analisis awal menunjukkan bahwa model 3 Parameter merupakan model yang paling baik dibandingkan dengan model 1 dan 2 parameter. Berdasarkan informasi dari tabel di atas, ditetapkan analisis data pada penelitian ini dilakukan dengan model 3 parameter.

Berdasarkan hasil analisis dengan *Bilog MG*, parameter butir perangkat soal UKK Bahasa Inggris 3 kabupaten di DIY dapat diketahui. Sebelum hasil analisis dapat dilihat pada *phase 3*, dilakukan analisis awal untuk melihat butir yang fit. Butir yang fit memiliki *output* korelasi biserialnya  $> 0,3$ . Hasil analisis awal ini menunjukkan bahwa terdapat 50 butir soal yang memiliki *output* korelasi biserial  $> 0,3$ . Dengan kata lain, terdapat 17 butir soal yang tidak fit karena memiliki *output* korelasi biserial  $< 0,3$ . Butir-butir soal tersebut antara lain butir soal nomor 6, 9, 15, 27, 52, 62, 76, 81, 83, 87, 106, 109, 135, 138, 141, 142, serta butir ke-149. Ke-17 butir soal tersebut tidak akan diikutkan dalam analisis berikutnya karena butir yang memiliki korelasi biserial  $< 0,3$  menyebabkan tidak munculnya *output phase 3* dari *Bilog MG*.

Berdasarkan hasil analisis terhadap ke 50 butir soal tersebut, tampak bahwa terdapat 37 butir soal (74%) yang memiliki daya beda, tingkat kesulitan maupun *pseudoguessing* yang memenuhi syarat sebagai butir soal yang baik. Berdasarkan parameter daya beda, seluruh 50 butir soal (100%) memiliki daya beda yang memenuhi syarat butir soal yang baik. Berdasarkan parameter tingkat kesulitan soal, seluruh 50 butir soal (100%) memiliki tingkat kesulitan yang memenuhi syarat butir soal yang baik. Berdasarkan parameter *pseudoguessing*, terdapat 13 butir soal (26%) yang kurang baik karena nilainya terlalu tinggi ( $> 0,25$ ). Pada hasil analisis atau *output phase 3* dengan *Bilog MG*, diperoleh informasi bahwa indeks reliabilitas ketiga tes tersebut adalah 0,912. Dari informasi tersebut, dapat dikatakan bahwa perangkat tes UKK Bahasa Inggris kelas VIII tahun 2009 dari Sleman, Gunungkidul, dan Kulonprogo reliabel.

Dari hasil analisis dengan *Bilog MG phase 3*, diperoleh juga informasi mengenai fungsi informasi butir yang menyatakan kekuatan atau sumbangan butir dalam mengungkap *latent trait* yang diukur oleh soal tersebut. Butir soal dengan fungsi informasi yang tinggi, memiliki kesalahan pengukuran yang kecil, sehingga memberikan kontribusi yang tinggi terhadap nilai fungsi informasi tes.

Rekap butir-butir soal tersebut, beserta dengan indikatornya dapat dilihat pada Tabel berikut.

Tabel 1. Rangkuman Butir Soal Yang Memuat Daya Beda, Tingkat Kesulitan, *Pseudoguessing*, Fungsi Informasi, *Theta*, Dan Indikator Butir Soal

Butir ke-	Daya Beda	Tingkat Kesulitan	<i>Pseudoguessing</i>	Fungsi Informasi Butir Maksimum	<i>Cocok untuk skala Theta (θ)</i>	Indikator
2	0,752	0,62	0,182	0,288	0,8	Mengidentifikasi makna gagasan (ideasional): informasi rinci (waktu) dari teks berbentuk undangan.
7	1,093	0,708	0,185	0,604	0,8	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks berbentuk naratif.
13	1,046	1,168	0,187	0,548	1,4	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks recount.
14	0,636	0,829	0,174	0,209	1,0	Mengidentifikasi makna gagasan (ideasional): informasi faktual dari teks naratif.
17	0,959	1,338	0,181	0,466	1,4	Mengidentifikasi makna gagasan (ideasional): informasi faktual dari teks berbentuk peringatan pendek.
18	0,812	1,453	0,135	0,367	1,6	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks recount.

Butir ke-	Daya Beda	Tingkat Kesulitan	<i>Pseudoguessing</i>	Fungsi Informasi Butir Maksimum	<i>Cocok untuk skala Theta (θ)</i>	<i>Indikator</i>
19	0,791	0,708	0,196	0,308	1,0	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks.
21	1,583	-0,759	0,072	1,550	-0,8	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks naratif.
29	1,27	1,098	0,261	0,286	1,6	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks naratif.
42	0,789	1,333	0,235	0,188	0,8	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks recount pendek.
43	0,966	1,212	0,4	0,208	1,2	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks berbentuk recount.
49	0,542	0,743	0,06	0,767	-0,4	Menyusun kalimat menjadi paragraph yang benar dan bermakna.
53	0,575	1,081	0,07	0,526	0,8	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks berbentuk pengumuman pendek.
55	1,113	-0,489	0,079	0,173	1,0	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks berbentuk recount.
57	0,941	0,645	0,099	0,565	0,2	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks berbentuk naratif.
59	0,574	0,725	0,164	0,181	0,6	Mengidentifikasi makna gagasan (ideasional): informasi rinci (waktu) dari teks pesan pendek (email) berbentuk email.

Butir ke-	Daya Beda	Tingkat Kesulitan	<i>Pseudoguessing</i>	Fungsi Informasi Butir Maksimum	<i>Cocok untuk skala Theta (θ)</i>	<i>Indikator</i>
61	0,863	0,12	0,35	0,285	1,4	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks pendek (email) berbentuk deskriptif.
64	1,108	0,07	0,234	0,497	0,8	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks berbentuk recount.
65	0,562	0,405	0,118	0,124	0,0	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks berbentuk recount.
68	0,789	1,201	0,236	0,176	-0,6	Mengidentifikasi makna gagasan (ideasional): informasi faktual dari teks berbentuk naratif.
69	0,718	-0,233	0,307	0,207	0,4	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks berbentuk naratif.
77	0,933	0,668	0,122	0,610	0,6	Mengidentifikasi makna gagasan (ideasional) tujuan komunikatif teks fungsional pendek berbentuk (prosedur) label obat.
82	0,495	-0,235	0,183	0,123	1,0	Menyusun kata menjadi kalimat secara urut dan benar.
88	0,596	-0,904	0,194	0,826	0,4	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks naratif.
91	0,595	0,338	0,106	0,102	0,6	Menyusun kalimat menjadi paragraf naratif secara urut dan benar.
93	0,726	0,588	0,406	0,398	0,0	Menyusun kalimat menjadi paragraf naratif secara urut dan benar.
95	0,992	0,481	0,078	0,288	1,6	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks berbentuk pengumuman singkat.

Butir ke-	Daya Beda	Tingkat Kesulitan	<i>Pseudoguessing</i>	Fungsi Informasi Butir Maksimum	<i>Cocok untuk skala Theta (θ)</i>	<i>Indikator</i>
100	0,516	0,695	0,231	0,771	1,2	Menyusun kalimat menjadi paragraf naratif secara urut dan benar.
108	0,846	0,994	0,278	0,199	1,8	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks berbentuk pengumuman.
116	1,222	0,243	0,136	0,270	1,0	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks naratif.
117	0,463	0,222	0,217	0,703	1,2	Mengidentifikasi tujuan komunikatif teks fungsional pendek.
125	0,672	0,029	0,4	0,340	1,4	Mengidentifikasi makna gagasan (ideasional): informasi rinci dari teks berbentuk pengumuman singkat.
127	0,892	-0,136	0,191	0,459	0,6	Mengidentifikasi makna gagasan (ideasional): gagasan utama dari teks berbentuk surat.
128	1,201	0,264	0,299	0,417	-0,2	Mengidentifikasi makna gagasan (ideasional): makna kata dari teks berbentuk surat.
146	0,788	1,31	0,229	0,529	0,4	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks naratif.
147	0,86	0,983	0,254	0,694	1,8	Mengidentifikasi makna gagasan (ideasional): mengisi titik-titik dalam kalimat yang berupa kata berdasarkan konteks dari teks naratif.
150	1,193	0,457	0,333	0,153	0,4	Menyusun kata menjadi kalimat secara urut dan benar.

Seperti telah dibahas sebelumnya, berdasarkan hasil analisis empiris dengan menggunakan program komputer *Bilog MG phase 3*, diketahui bahwa 37 butir soal tergolong dalam kategori baik berdasarkan kriteria daya beda, tingkat kesulitan dan juga *pseudoguessing*. Butir-butir tersebut adalah butir soal nomor, 2, 7, 13, 14, 17, 18, 19, 21, 29, 42, 43, 49, 53, 55, 57, 59, 61, 64, 65, 68, 69, 77, 82, 88, 91, 93, 95, 100, 108, 116, 117, 125, 127, 128, 146, 147, serta butir soal 150.

Setelah butir-butir tersebut diketahui karakteristiknya baik berdasarkan hasil analisis kualitatif dengan *expert judgment*, analisis kuantitatif dengan *Iteman 3.00* dan analisis dengan *Bilog MG. 3.0*, serta diketahui fungsi informasi butir dan *theta*-nya, ke-37 butir soal tersebut kemudian dapat dimasukkan dalam rintisan bank soal. Butir-butir yang sudah masuk dalam rintisan bank soal memuat informasi mengenai jenis sekolah, standar kompetensi, kompetensi dasar, indikator, daya beda, tingkat kesulitan butir, *pseudoguessing* serta fungsi informasi butir. Fungsi informasi butir dan skala *theta* dapat digunakan sebagai acuan dalam menentukan peserta didik dengan kemampuan seberapa yang cocok di tes dengan butir soal tersebut. Selain itu, fungsi informasi dan skala *theta* juga dapat digunakan untuk menyusun perangkat soal bagi kelompok siswa tertentu, semisal kelompok siswa dengan kemampuan tinggi, dengan cara mengambil butir soal dengan skala *theta* yang tinggi. Salah satu contoh butir dalam rintisan bank soal dapat dilihat pada gambar berikut:

<b>KARTU SOAL BENTUK PILIHAN GANDA</b>						
Jenis Sekolah : SMP		Penyusun : MKKS GK				
Mata Pelajaran : Bahasa Inggris		Tahun Ajaran : 2008/2009				
Bahan Kelas/smt : VIII/2						
Bentuk Tes : Tertulis						
<p><b>Standar Kompetensi:</b> 2. Berkomunikasi secara lisan dan tertulis dengan menggunakan ragam bahasa yang sesuai dengan lancar dan akurat dalam wacana interaksional dan atau monolog terutama dalam wacana berbentuk naratif, recount, deskriptif, dan anekdot.</p>	No Soal	Kunci	Kode Soal <b>GK 01</b>			
	2	C				
<b>Materi</b> <i>Inviting/mengundang</i>	<p style="text-align: center;"><b>Rumusan Butir Soal:</b></p> <div style="border: 1px solid black; padding: 10px; margin: 10px auto; width: 80%;"> <p style="text-align: center;">Dear Clara, Come and join us at a ( surprise ) birthday party for Amanda ! Let's eat, drink and be merry ! Celebrations will be held at Roro Jonggrang Room, Saphir Hotel on Wednesday July 9 at 6.30 PM Look forward to seeing you... R.S.V.P. call : Shinta 081. 768. 2603</p> </div> <p>2. When will the party be held ?</p> <p>a. In the morning.                      c. In the evening. b. In the afternoon                    d. At night</p>					
<b>Indikator</b> Mengidentifikasi makna gagasan (ideasional): informasi rinci (waktu) dari teks berbentuk undangan.						
No	Digunakan untuk	Daya Beda	Tingkat Kesulitan	<i>Pseudoguessing</i>	Fungsi informasi	<i>Cocok untuk skala Theta (θ) (Kemampuan)</i>
2		0,752	0,62	0,182	0,3	0,8

Contoh salah satu butir yang sudah disusun menjadi rintisan bank soal.

Gambar di atas menunjukkan bahwa butir soal 2 merupakan contoh butir soal yang dapat digunakan untuk menguji kemampuan siswa dalam mengundang/*inviting* dengan indikator mengidentifikasi makna gagasan (*ideasional*): informasi rinci (waktu) dari teks berbentuk undangan. Butir tersebut memiliki daya beda yang baik, yakni 0,752, tingkat kesulitan yang sedang yakni 0,62, kemungkinan menebak jawaban dengan benar sebesar 0,182, fungsi informasi 0,3 serta dapat digunakan untuk mengukur hasil belajar pada siswa dengan kemampuan menengah atau sedang yakni pada *theta* 0,8.

Selanjutnya, bank soal yang akan disusun hendaknya dapat mengikuti bentuk rintisan model seperti dalam penelitian ini, sehingga nantinya bank soal yang disusun dapat memberikan informasi yang cukup.

## **Simpulan**

Berdasarkan hasil penjabaran pada bab sebelumnya ditemukan beberapa hal sebagai berikut.

1. Hasil analisis secara kualitatif dengan *expert judgment* menunjukkan bahwa dari 150 butir soal perangkat tes UKK Bahasa Inggris SMP kelas VIII di 3 kabupaten di DIY, sebanyak 106 butir soal (70,66%) dinyatakan baik dan sebanyak 44 butir soal (29,33%) tergolong “kurang baik”.
2. Hasil analisis secara kuantitatif dengan program *Iteman 3.00* menunjukkan bahwa 67 (63,21%) butir soal dari 3 kabupaten tersebut tergolong baik, sedangkan 39 butir lainnya (36,79%) tergolong kurang baik.
3. Analisis dengan *Bilog MG 3 PL* menunjukkan bahwa 50 butir soal dari 67 butir tergolong fit; 37 butir soal (74%) tergolong dalam kategori baik sedangkan 13 butir lainnya (26%) tergolong kurang baik.
4. Sistem pengujian UKK Bahasa Inggris SMP kelas VIII di Kabupaten Sleman, Gunungkidul, dan Kulonprogo belum menggunakan bank soal yang memuat informasi tentang tingkat kesulitan, daya beda, *pseudoguessing*, SK, KD, dan indikator sebagai sumber pengambilan butir soal sehingga dari hasil analisis, baik analisis kualitatif dengan *expert*

*judgment* maupun analisis kuantitatif dengan *Iteman 3.00* serta *Bilog MG*, jumlah butir soal yang kurang baik di tiap kabupaten masih cukup banyak.

### **Implikasi**

1. Perlu dilakukan pelatihan analisis butir soal baik dengan teori tes klasik maupun teori modern untuk guru. Hal ini perlu dilakukan untuk membekali guru dengan kemampuan minimal menganalisis perangkat tes, paling tidak dengan teori klasik sehingga guru dapat mengaplikasikannya. Dengan kemampuan menganalisis butir soal dengan teori tes klasik, guru dapat langsung mengetahui karakter perangkat tes yang mereka susun maupun mereka gunakan sehingga mereka dapat membantu meningkatkan kualitas evaluasi hasil pembelajaran.
2. Perlu diberikan pelatihan penyusunan rintisan bank soal Bahasa Inggris berbasis provinsi kepada guru-guru Bahasa Inggris SMP di DIY. Rintisan bank soal tersebut dapat disusun dari butir-butir soal dari perangkat soal yang sudah ada. Penyusunan tersebut dapat dilakukan dengan mengikuti langkah-langkah sebagai berikut.
  - a. Pengumpulan perangkat soal serta respon siswa yang tertuang dalam lembar jawab. Respon siswa tersebut kemudian disusun dalam *Editplus 2* untuk memudahkan pengolahan pada tahap selanjutnya.
  - b. Telaah butir soal untuk melihat apakah butir soal sudah sesuai dengan indikator pencapaian hasil belajar yang disesuaikan dengan SK, dan KD.
  - c. Telaah butir soal secara kualitatif dengan *expert judgment* untuk melihat butir soal yang baik berdasarkan kriteria konstruksi, bahasa, dan materi.
  - d. Hasil telaah butir secara kualitatif digunakan sebagai dasar untuk telaah butir secara kuantitatif dengan *Iteman 3.00* yakni untuk melihat karakteristik butir soal secara tradisional dengan berdasarkan kriteria tingkat kesulitan, daya beda, dan proporsi pemilih atau distraktor.
  - e. Hasil telaah butir dengan *Iteman 3.00* digunakan sebagai dasar untuk telaah butir secara kuantitatif dengan *Bilog MG* untuk melihat butir

yang fit serta karakteristik butir soal pilihan ganda yang baik berdasarkan kriteria tingkat kesulitan, daya beda serta *psudoguessing* atau kemungkinan menebak jawaban dengan benar.

### **Keterbatasan Penelitian**

Penelitian ini pada dasarnya bertujuan untuk menyusun rintisan bank soal Bahasa Inggris SMP kelas VIII berbasis provinsi, namun demikian terdapat keterbatasan dalam penelitian ini yakni data yang berupa perangkat soal dan hasil jawaban siswa dalam UKK Bahasa Inggris SMP kelas VIII tahun 2009 hanya dikumpulkan dari 3 kabupaten di DIY yakni Kabupaten Gunungkidul, Sleman, dan Yogyakarta. Pada awalnya, penelitian ini direncanakan untuk menggunakan data dari 5 kabupaten di DIY, namun demikian pada pelaksanaannya, peneliti menemukan banyak kesulitan dalam mengumpulkan perangkat soal beserta jawaban dari Kabupaten Bantul dan Kotamadya Yogyakarta. Keterbatasan ini disebabkan sulitnya mengkoordinasikan pengumpulan soal dari Bantul dan Kodya karena tidak ada data yang sudah berupa *soft copy* hasil UKK dari kabupaten dan kodya tersebut. Meskipun peneliti telah menyampaikan surat pengantar permohonan perangkat soal dan hasil UKK Bahasa Inggris SMP, namun Dinas Pendidikan Kabupaten Bantul dan Kodya Yogyakarta tidak dapat membantu pengumpulan data tersebut. Dengan demikian, data yang digunakan hanya data dari Kabupaten Gunungkidul, Sleman, dan Yogyakarta.

### **Saran**

1. Pengujian hasil belajar siswa pada mata pelajaran Bahasa Inggris hendaknya tidak hanya mengukur kemampuan membaca saja, tetapi juga kemampuan berbicara, menyimak, serta menulis.
2. Sebelum perangkat soal digunakan untuk pengujian hasil belajar siswa, terutama dalam mata pelajaran Bahasa Inggris, perlu dilakukan analisis kualitatif dengan *expert judgement*. Teknik ini dapat dilakukan dengan meminta ahli yang sudah memiliki pengalaman dalam hal pembelajaran

Bahasa Inggris serta penilaian. Analisis kualitatif dengan *expert judgment* dilakukan untuk mengetahui karakteristik soal dengan melihat kriteria konstruksi, bahasa serta materi.

3. Selain analisis dengan *expert judgement*, analisis empiris dengan teori tes klasik (*Iteman 3.00*) juga diperlukan untuk melihat karakteristik butir soal dilihat dari kriteria tingkat kesulitan, daya beda, serta proporsi pemilih.
4. Setelah di analisis empiris dengan teori klasik, diperlukan juga analisis dengan teori respon butir atau teori tes modern untuk melihat karakteristik butir soal berdasarkan kriteria daya beda, tingkat kesukaran tes maupun *pseudoguessing* serta tingkat kemampuan peserta tes, agar interpretasi terhadap hasil tes menghasilkan keputusan yang dapat dipertanggungjawabkan. Perhitungan nilai informasi tes serta SEM perlu dilakukan untuk mengetahui besarnya fungsi informasi tes dan kesalahan pengukuran yang dilakukan dari tes yang digunakan, serta untuk mengetahui tingkat efisiensi dan kecocokan paket tes yang diberikan dengan tingkat kemampuan peserta.
5. Perlu disusun bank soal berbasis provinsi untuk mata pelajaran Bahasa Inggris serta mata pelajaran lain pada umumnya untuk membantu guru dalam melakukan pengukuran hasil belajar siswa. Keberadaan rintisan bank soal Bahasa Inggris serta mata pelajaran lain juga sangat diperlukan untuk membantu pemerintah dalam menentukan mutu pendidikan sehingga dari hasil penilaian pembelajaran dapat digunakan sebagai pedoman perbandingan mutu pendidikan antar kabupaten.

### **Daftar Pustaka**

- Bachman, L.F. and Palmer, A.S. (1996). *Language testing in practice*. Oxford University press: Oxford.
- Chang, Chih-Chung. (2005). Developing tailored instruments: item banking and computerized adaptive assessment. *Evaluation and program planning journal*, 24, 251-215.

- Depdikbud. (1995). *Petunjuk Pelaksanaan pengembangan bank soal di wilayah*. Balitbang: Jakarta.
- Dikmenum. (2006). *Panduan analisis butir soal*. Diambil pada 21 Juli 2009, dari <http://www.dikmenum.go.id/dataapp/kurikulum/5.%20PE-RANGKAT%20PENILAIAN%20KTSP%20SMA/6.%20ANALISIS%20BUTIR%20SOAL/PANDUAN%20ANALISIS%20BUTIR%20SOAL.doc>.
- Eid, G. K. (2005). *The effects of sample size on the equating test items*. Proquest Education Journals. (165-180). Kuwait University.
- Fernandes, H.J.X. (1984). *Testing and measurement*. National Education Planning, Evaluation and Curriculum Development: Jakarta.
- Mardapi, Djemari. (2008). *Teknik penyusunan instrumen tes dan nontes*. Mitra Cendikia: Yogyakarta.
- Wright, B.D. and Mark H.S.(1979). *Best test design*. Mesa Press. Chicago.