

ANALISIS BUTIR SOAL DENGAN ANBUSO¹

Oleh: Ali Muhson, Kiromim Baroroh dan Mustofa²

A. Pendahuluan

Untuk mendapatkan instrumen berkualitas tinggi, selain dilakukan analisis secara teori (telaah butir berdasarkan aspek isi, konstruksi, dan bahasa) perlu juga dilakukan analisis butir secara empirik. Secara garis besar, analisis butir secara empirik ini dapat dibedakan menjadi dua, yaitu dengan pendekatan teori tes klasik (*Classical Test Theory*) dan teori respons butir (*Item Response Theory, IRT*).

Teori tes klasik atau disebut teori skor murni klasik (Allen & Yen, 1979:57) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Jika dituliskan dengan pernyataan matematis, maka kalimat tersebut menjadi

$$X = T + E$$

dengan :

X : skor amatan,

T : skor sebenarnya,

E : skor kesalahan pengukuran (*error score*).

Kesalahan pengukuran yang dimaksudkan dalam teori ini merupakan kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan penyimpangan secara teoritis dari skor amatan yang diperoleh dengan skor amatan yang diharapkan. Kesalahan pengukuran yang sistematis dianggap bukan merupakan kesalahan pengukuran.

Ada beberapa asumsi dalam teori tes klasik. Skor kesalahan pengukuran tidak berinteraksi dengan skor sebenarnya, merupakan asumsi yang pertama. Asumsi yang kedua

¹ Disampaikan dalam Kegiatan Pengabdian pada Masyarakat berupa “Pelatihan Analisis Butir Soal dan Program Remedial dengan Software AnBuso (Strategi Meningkatkan Kemampuan Guru Ekonomi Dalam Mengembangkan Tes Berkualitas)” di FE UNY pada tanggal 12-13 Juli 2012

² Staf Pengajar Jurusan Pendidikan Ekonomi FE UNY

adalah skor kesalahan tidak berkorelasi dengan skor sebenarnya dan skor-skor kesalahan pada tes-tes yang lain untuk peserta tes (*testee*) yang sama. Ketiga, rata-rata dari skor kesalahan ini sama dengan nol. Asumsi-asumsi pada teori tes klasik ini dijadikan dasar untuk mengembangkan formula-formula dalam menentukan validitas dan reliabilitas tes.

Validitas dan reliabilitas pada perangkat tes digunakan untuk menentukan kualitas tes. Kriteria lain yang dapat digunakan untuk menentukan kualitas tes adalah indeks kesukaran, daya pembeda dan efektivitas distraktor.

B. Reliabilitas

Mehrens & Lehmann (1973: 102) menyatakan bahwa reliabilitas merupakan derajat keajegan (*consistency*) di antara dua buah hasil pengukuran pada objek yang sama. Definisi ini dapat diilustrasikan dengan seseorang yang diukur tinggi badannya akan diperoleh hasil yang tidak berubah walaupun menggunakan alat pengukur yang berbeda dan skala yang berbeda. Dalam kaitannya dengan dunia pendidikan, prestasi atau kemampuan seorang siswa dikatakan reliabel jika dilakukan pengukuran, hasil pengukuran akan sama informasinya, walaupun penguji berbeda, korektornya berbeda atau butir soal yang berbeda tetapi memiliki karakteristik yang sama.

Allen & Yen (1979: 62) menyatakan bahwa tes dikatakan reliabel jika skor amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Selanjutnya dinyatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Dengan demikian, pengertian yang dapat diperoleh dari pernyataan tersebut adalah suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya.

Dalam pendidikan, pengukuran tidak dapat langsung dilakukan pada ciri atau karakter yang akan diukur. Ciri/karakter ini bersifat abstrak. Hal ini menyebabkan sulitnya memperoleh alat ukur yang stabil untuk mengukur karakteristik seseorang (Mehrens & Lehmann, 1973: 103).

Berdasarkan uraian di atas, maka dalam pembuatan alat ukur dalam dunia pendidikan harus dilakukan secermat mungkin dan disesuaikan dengan kaidah-kaidah yang telah ditentukan oleh ahli-ahli pengukuran di bidang pendidikan. Untuk melihat reliabilitas suatu alat

ukur, yang berupa suatu indeks reliabilitas, dapat dilakukan penelaahan secara statistik. Nilai ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*).

Untuk menentukan nilai reliabilitas suatu tes (butir soal berbentuk pilihan ganda (*multiple choice*)) dapat digunakan formula sebagai berikut

$$\hat{\alpha} = \frac{R}{R-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

dengan :

R : banyaknya butir soal,

σ^2 : varians.

Mehrens & Lehmann (1973: 104) menyatakan bahwa meskipun tidak ada perjanjian secara umum, tetapi secara luas dapat diterima bahwa untuk tes yang digunakan untuk membuat keputusan pada siswa secara perorangan harus memiliki koefisien reliabilitas minimal sebesar 0,85. Dengan demikian, pada penelitian ini, tes seleksi digunakan untuk menentukan keputusan pada siswa secara perorangan, sehingga indeks koefisien reliabilitasnya diharapkan minimal sebesar 0,85.

C. Validitas

Validitas suatu perangkat tes dapat diartikan merupakan kemampuan suatu tes untuk mengukur apa yang seharusnya diukur (Azwar, 2012: 51). Menurut Cohen-Swerdlik (2009) *validity, as applied to a test, is a judgment or estimate of how well a test measures what it purports to measure in a particular context*. Pengertian ini menunjukkan bahwa validitas dalam sebuah tes menjadi hal yang sangat penting karena akan mengukur kemampuan peserta didik secara tepat.

Validitas juga merujuk pada ketepatan (*appropriateness*), kebermaknaan (*meaningfulness*) dan kemanfaatan (*usefulness*) kesimpulan yang didapatkan dari interpretasi skor tes (Kusaeri dan Suprananto, 2012). Ada tiga tipe validitas, yaitu validitas isi, validitas konstruk dan validitas kriteria (Cohen-Swerdlik 2009: 185; Azwar, 2012: 52).

Ada dua macam validitas isi , yaitu validitas kenampakan dan validitas logika (Azwar, 2012). Validitas isi berarti sejauh mana suatu perangkat tes mencerminkan keseluruhan kemampuan yang hendak diukur (Azwar, 2012), yang berupa analisis rasional terhadap domain yang hendak diukur. Validitas kenampakan didasarkan pada pertanyaan apakah suatu butir-butir dalam perangkat tes mengukur aspek yang relevan dengan domainnya. Validitas logika berkaitan dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah keseluruhan butir merupakan sampel representatif dari keseluruhan butir yang mungkin dibuat.

Validitas kriteria, disebut juga validitas prediktif, merupakan kesahihan suatu perangkat tes dalam membuat prediksi, dapat meramalkan keberhasilan siswa pada masa yang akan datang. Validitas prediktif suatu perangkat tes dapat diketahui dari korelasi antara perangkat tes dengan kriteria tertentu yang dikehendaki, yang disebut dengan variabel kriteria (Cohen-Swerdlik, 2009: 97; Azwar, 2012).

D. Tingkat Kesukaran

Tingkat kesukaran suatu butir soal, yang disimbolkan dengan p_i , merupakan salah satu parameter butir soal yang sangat berguna dalam penganalisan suatu tes. Hal ini disebabkan karena dengan melihat parameter butir ini, akan diketahui seberapa baiknya kualitas suatu butir soal. Jika p_i mendekati 0, maka soal tersebut terlalu sukar, sedangkan jika p_i mendekati 1, maka soal tersebut terlalu mudah, sehingga perlu dibuang. Hal ini disebabkan karena butir tersebut tidak dapat membedakan kemampuan seorang siswa dengan siswa lainnya.

Sudjana (2011: 137) menyatakan bahwa secara umum indeks kesukaran suatu butir sebaiknya terletak dalam kategori sedang yakni pada interval 0,31 – 0,70. Pada interval ini, informasi tentang kemampuan siswa akan diperoleh secara maksimal. Dalam merancang indeks kesukaran suatu perangkat tes, perlu dipertimbangkan tujuan penyusunan perangkat tes tersebut. Untuk menentukan indeks kesukaran dari suatu butir pada perangkat tes pilihan ganda, digunakan persamaan sebagai berikut:

$$p_i = \frac{\sum B}{N}$$

dengan:

p = proporsi menjawab benar pada butir soal tertentu.

ΣB = banyaknya peserta tes yang menjawab benar.

N = jumlah peserta tes yang menjawab.

E. Daya Pembeda

Untuk menentukan daya pembeda, dapat digunakan indeks diskriminasi, indeks korelasi biserial, indeks korelasi *point biserial*, dan indeks keselarasan. Pada analisis butir dalam penelitian ini, hanya digunakan indeks korelasi *point biserial*. Koefisien korelasinya untuk suatu butir tes ditentukan dengan rumus:

$$r_{pbis} = \left[\frac{\bar{X}_1 - \bar{X}}{s_x} \right] \sqrt{\frac{p_1}{1 - p_1}}$$

dengan r_{pbis} = koefisien korelasi point biserial, X_i merupakan variabel kontinyu, \bar{X}_1 merupakan rerata skor X untuk peserta tes yang menjawab benar butir tersebut, \bar{X} merupakan rerata skor X , s_x merupakan standar deviasi dari skor X , dan p_1 merupakan proporsi peserta tes yang menjawab benar butir tersebut.

Pada suatu butir soal, indeks daya beda dikatakan baik jika lebih besar atau sama dengan 0,3 (Nunnally & Bernstein, 2009: 304; Ebel & Frisbie, 1991; 232). Indeks daya pembeda suatu butir yang kecil nilainya akan menyebabkan butir tersebut tidak dapat membedakan siswa yang kemampuannya tinggi dan siswa yang kemampuannya rendah. Pada analisis tes dengan *Content-Referenced Measures*, indeks daya pembeda butir tidak terlalu perlu menjadi perhatian, asalkan tidak negatif (Ebel & Frisbie, 1991). Jika nilainya kecil, menunjukkan bahwa kemencengan distribusi skor dari populasi, yang juga mengakibatkan validitas tes menjadi rendah.

F. Efektivitas Distraktor

Menurut Sudijono (2012) pada saat membicarakan tes objektif bentuk multiple choice *item* tersebut untuk setiap butir *item* yang dikeluarkan dalam tes hasil belajar telah dilengkapi dengan beberapa kemungkinan jawab, atau yang sering dikenal dengan istilah option atau alternatif.

Option atau alternatif itu jumlahnya berkisar antara 3 sampai dengan 5 buah, dan dari kemungkinan-kemungkinan jawaban yang terpasang pada setiap butir *item* itu, salah satu di antaranya adalah merupakan jawaban betul (kunci jawaban), sedangkan sisanya adalah merupakan jawaban salah. Jawaban-jawaban salah itulah yang biasa dikenal dengan istilah *distractor* (pengecoh).

Menganalisis fungsi distraktor sering dikenal dengan istilah lain, yaitu : menganalisis pola penyebaran jawaban *item*. Adapun yang dimaksud dengan pola penyebaran jawaban *item* adalah suatu pola yang dapat menggambarkan bagaimana *testee* menentukan pilihan jawabnya terhadap kemungkinan-kemungkinan jawab yang telah dipasangkan pada setiap butir *item*.

Suatu kemungkinan dapat terjadi, yaitu bahwa dari keseluruhan alternatif yang dipasang pada butir *item* tertentu, sama sekali tidak dipilih oleh *testee*. Dengan kata lain, *testee* menyatakan “blangko”. Pernyataan blangko ini sering dikenal dengan istilah omiet dan biasa diberi lambang dengan huruf O.

Distraktor dinyatakan telah dapat menjalankan fungsinya dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih oleh 5 % dari seluruh peserta tes.

Sebagai tindak lanjut atas hasil penganalisaan terhadap fungsi distraktor tersebut maka distraktor yang sudah dapat menjalankan fungsinya dengan baik dapat dipakai lagi pada tes-tes yang akan datang, sedangkan distraktor yang belum dapat berfungsi dengan baik sebaiknya diperbaiki atau diganti dengan distraktor yang lain.

G. Analisis Butir Soal dengan AnBuso

Software AnBuso (Analisis Butir Soal) merupakan program analisis butir soal yang dikembangkan secara sederhana untuk membantu dalam membuat administrasi guru, khususnya yang terkait dengan analisis butir soal. AnBuso dikembangkan dengan program Microsoft Excel sehingga mempermudah guru dalam menggunakannya

Ada beberapa alasan mengapa perlu menggunakan AnBuso, di antaranya adalah:

- Sederhana programnya
- Mudah menggunakannya
- Compatible

- Praktis penggunaannya
- Tersedia juga untuk tes subjektif
- Ada pengelompokan remedial
- Hasil analisis dalam format laporan
- Hasil analisis grafik tersedia
- Tidak berbayar (*Free Charge*)

Di dalam AnBuso berisi dua hal yaitu data input dan laporan. Data input terdiri dari **Input 01** yang memberi kesempatan untuk mengisi data umum, dan **Input 02** yang berisi identitas peserta tes dan jawabannya. Jika kedua input tersebut sudah terisi maka secara otomatis akan dibuat laporan yang meliputi:

- Laporan Peserta
- Laporan Butir
- Pola Jawaban Butir
- Laporan Essay
- Materi Remedial
- Peserta Remedial
- Grafik

Data yang perlu dimasukkan dalam **Input 01** meliputi:

1. Satuan Pendidikan
2. Mata Pelajaran
3. Kelas/Program
4. Nama Tes SK/KD
5. Nama Guru
6. NIP
7. Semester
8. Tahun Pelajaran
9. Tanggal Tes
10. Tanggal Diperiksa
11. Nama Kepala Sekolah

12. NIP Kepala Sekolah

13. Tempat Laporan

14. Tanggal Laporan

15. Skala Penilaian (10 atau 100)

16. Nilai KKM

17. Data soal objektif yang meliputi:

- Jumlah Alternatif Jawaban:
 - Isikan 4 jika alternatif jawabannya A, B, C, dan D
 - Isikan 5 jika alternatif jawabannya A, B, C, D, dan E
 - Dan sebagainya
- Skor Benar tiap Butir Soal
 - Isikan skor yang diperoleh peserta tes jika ia menjawab dengan benar pada setiap butir soal tersebut
- Skor Salah tiap butir soal
 - Isikan skor yang diperoleh peserta tes jika ia menjawab salah pada setiap butir soal tersebut
- Kunci Jawaban
 - Isikan kunci jawaban soal secara berurutan mulai nomor satu sampai yang terakhir dengan format menggunakan huruf besar tanpa spasi. Misalnya DDADDCADBDDDBCABDDDD
 - Jumlah soal maksimal 50 sehingga dalam penulisan kunci jawaban maksimal 50 karakter
- Kompetensi Dasar
 - Isikan kompetensi dasar yang diukur dalam setiap butir soal. Isian ini diperlukan untuk menganalisis KD mana yang belum dikuasai peserta tes guna merancang program remedial baik klasikal maupun individual

18. Data Soal Essay yang perlu diisikan meliputi:

- Jumlah Soal (maksimal 10):

- Isikan jumlah soal essay yang akan dianalisis. Jumlah soal yang disediakan dalam program ini maksimal 10 soal
 - Skor Maksimal Soal
 - Isikan skor maksimal yang dapat diperoleh peserta tes pada setiap butir soalnya. Skor tersebut bisa berbeda-beda untuk setiap soalnya.
 - Kompetensi Dasar
 - Isikan kompetensi dasar yang diukur dalam setiap butir soal.
- Data yang perlu dimasukkan dalam **Input 02** meliputi:
- Nama Peserta Tes
 - Jenis Kelamin (L = Laki-laki, P = Perempuan)
 - Jawaban Soal Pilihan Ganda
 - Isikan jawaban peserta tes secara berurutan mulai nomor satu sampai yang terakhir dengan format menggunakan huruf besar tanpa spasi. Misalnya DDADDCADBDDDBCABDDDD
 - Untuk jawaban kosong atau butir pertanyaan yang tidak bisa dijawab peserta tes dapat diisikan spasi atau karakter lain selain alternatif jawaban
 - Skor Jawaban Essay
 - Isikan skor yang diperoleh peserta tes untuk masing-masing butir soal.
- Jika semua data yang ada dalam **Input 01** dan **Input 02** sudah terisi maka program akan secara otomatis mengeluarkan laporan lengkap yang terdiri dari:
1. Laporan Peserta, berupa Daftar Nilai Ujian yang berisi tentang:
 - Daftar peserta dan jenis kelaminnya
 - Jumlah butir soal yang dijawab benar dan salah
 - Skor dan nilai tes objektif
 - Nilai tes essay
 - Nilai Akhir, dan
 - Keterangan (Lulus dan tidak lulus)

- Keterangan lain yang terdiri dari jumlah peserta tes, Jumlah (persentase) peserta yang lulus dan tidak lulus, serta nilai terendah, nilai tertinggi, nilai rata-rata dan standar deviasi
2. Laporan Butir, berupa hasil Analisis Butir Soal Pilihan Ganda yang berisi tentang:
- Daya Beda butir soal
 - Tingkat Kesukaran
 - Alternatif Jawaban tidak Efektif
 - Kesimpulan Akhir
3. Laporan Pola Jawaban Butir
- Laporan tentang Pola Jawaban Butir berisi tentang Sebaran Jawaban Soal Pilihan Ganda yang berupa persentase peserta tes yang menjawab alternatif jawaban yang tersedia.
4. Laporan Essay
- Laporan essay berupa Hasil Analisis Soal Essay yang berisi:
- Daya Beda butir soal
 - Kriteria sama dengan soal objektif
 - Tingkat Kesukaran
 - Kriteria sama dengan soal objektif
 - Kesimpulan Akhir
 - Baik jika daya beda baik/cukup baik dan tingkat kesukaran sedang
 - Cukup Baik jika salah satu di antara daya beda dan tingkat kesukaran tidak memenuhi syarat
 - Tidak baik jika daya beda dan tingkat kesukaran tidak memenuhi persyaratan
5. Laporan Materi Remedial
- Berupa Laporan Materi Remedial Individual dan Klasikal yang berisi tentang:
- Kompetensi dasar/materi yang tidak dikuasai secara individual
 - Kompetensi dasar/materi yang tidak dikuasai secara klasikal. KD dianggap tidak dikuasai secara klasikal jika peserta tes yang mampu menjawab dengan benar kurang dari 15%.
6. Laporan Peserta Remedial

Berupa Laporan Pengelompokan Peserta Remedial menurut Kompetensi Dasar/Materinya. Dalam laporan tersebut terlihat daftar peserta yang belum menguasai KD/Materi tertentu.

7. Grafik

- Grafik distribusi nilai dan KKM
- Pie Chart proporsi ketuntasan belajar

DAFTAR PUSTAKA

Azwar, Syaifuddin (2012) *Dasar-dasar Psikometri*. Yogyakarta: Pustaka Pelajar

Daryanto (2005) *Evaluasi Pendidikan*. Jakarta: Rineka Cipta

Depdiknas (2004). *Pedoman Umum Pengembangan Penilaian Kurikulum Berbasis Kompetensi SMA*. Jakarta: Direktorat Pendidikan Dasar dan Menengah, Direktorat Pendidikan Lanjutan Pertama.

Djaali & Mulyono, Pudji. (2007). *Pengukuran dalam Bidang Pendidikan*. Jakarta: Grasindo

Ebel, Robert L. & David A. Frisbie (1991) *Essential Of Educational Measurement (5th Edition)*. New Delhi: Prentice-Hall, Inc.

Kusaeri dan Suprananto (2012) *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu

Nunnally, Jum C. & Ira H. Bernstein (1994) *Psychometric Theory (3rd Edition)*. New York: McGraw-Hill, Inc.

Purwanti, Endang. (2008). *Asesmen Pembelajaran SD*. Direktorat Jendral Pendidikan Tinggi Departemen Pendidikan Nasional.

Puskur (2008) *Model Penilaian Kelas Kurikulum Berbasis Kompetensi*. Jakarta: Balitbang Depdiknas

Slameto (2001) *Evaluasi Pendidikan*. Jakarta: Bumi Aksara

Sudijono, Anas (2011) *Pengantar Evaluasi Pendidikan*. Jakarta: RajaGrafindo Persada

Sudjana, Nana (2011) *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya

Sukardi (2011) *Evaluasi Pendidikan: Prinsip & Operasionalnya*. Jakarta: Bumi Aksara

Widoyoko, Eko Putro (2009) *Evaluasi Program Pembelajaran*. Diambil dari <http://www.umpwr.ac.id/web/download/publikasi-ilmiah> pada tanggal 22 Maret 2012