PENGANTAR TEORI TES KLASIK (TTK)*)

Oleh :
Badrun Kartowagiran)**

KERJASAMA
PASCASARJANA UNY
DINAS PENDIDIKAN PROV DIY
2009

==========

*) Makalah disampaikan pada Pelatihan penulisan analisis butir dengan pendekatan TTK dan TRB tanggal 11 – 12 April 2009 di Lemlit UNY

**) Dosen Fakultas Teknik dan Pascasarjana UNY

PENGANTAR TEORI TES KLASIK

Oleh: Badrun Kartowagiran

1. Pendahuluan

Soal tes merupakan salah satu alat ukur yang sangat penting artinya bagi keakuratan informasi yang terkait dengan tingkat penguasaan siswa terhadap suatu materi pelajaran. Oleh karena itu soal tes harus disusun dengan sungguh-sungguh dan secermat mungkin. Menurut tujuan tes, penyusunan soal dapat dibagi menjadi dua, yaitu : (a) penyusunan soal tes umum, yakni soal yang tidak direncanakan untuk tujuan tertentu dan tidak memperhatikan karakteristik peserta, dan (b) penyusunan soal tes khusus yang direncanakan untuk tujuan khusus, misal tes kecepatan, dan lain sebagainya (Nunnaly, 1978). Penyusunan butir-butir soal di dalam penelitian ini termasuk dalam penyusunan soal tes umum.

2. Penyusunan Soal Tes Umum

Untuk menghasilkan soal tes berkualitas tinggi maka soal tes harus dikembangkan dengan cara yang sebaik-baiknya. Menurut Tim Pusisjian (1997/1998), langkah-langkah pengembangan suatu tes prestasi belajar adalah : (1) penentuan tujuan tes, (2) penyusunan kisi-kisi, (3) penulisan soal, (4) penelaahan soal (review dan revisi soal), (5) uji coba soal, termasuk analisis dan perbaikan, dan (6) perakitan soal menjadi perangkat tes.

a. Penentuan tujuan

Dalam melakukan pengetesan pasti ada tujuan yang ingin dicapai. Tujuan ini dapat berupa tujuan khusus, yaitu untuk melihat tingkat pencapaian suatu program.

Dalam dunia pendidikan, salah satu tujuan pengetesan adalah untuk mengetahui penguasaan peserta didik pada kompetensi/sub kompetensi tertentu setelah diajarkan. Penguasaan ini dapat diartikan, sejauh mana peserta didik memahami atau mungkin menganalisis materi tertentu yang telah dibahas di ruang kelas. Dapat pula tes tersebut digunakan untuk mengetahui kesulitan belajar peserta didik (diagnostik tes). Tujuan tes harus jelas agar arah dan ruang lingkup pengembangan tes selanjutnya juga jelas.

b. Penyusunan Kisi-kisi

Kisi-kisi tes yang juga disebut dengan *blue print* atau *table of spesification* diperlukan sebelum penyusunan soal tes dimulai. Kisi-kisi soal tes adalah deskripsi mengenai ruang lingkup dan isi dari materi yang akan diujikan, serta memberikan perincian mengenai soal-soal yang diperlukan oleh tes tersebut. Hal yang harus diperhatikan dalam menyusun kisi-kisi adalah indikator jabaran dari kempetensi dasar (KD), kompetensi dasar jabaran dari standar kompetensi (SK), standar kompetensi jabaran dari standar kompetensi lulusan mata pelajaran (SKL-MP), dan standar kompetensi lulusan mata pelajaran jabaran dari standar kompetensi lulusan satuan pendidikan (SKL-P), dan standar kompetensi lulusan satuan pendidikan jabaran dari Tujuan Pendidikan Nasional.

c. Penulisan butir-butir soal/tes

Penulisan butir-butir soal merupakan langkah penting dalam upaya pengembangan alat ukur kemampuan atau tes yang baik. Penulisan soal adalah penjabaran indikator jenis dan tingkat perilaku yang hendak diukur menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan perinciannya dalam kisi-kisi. Butir soal merupakan jabaran atau dapat juga ujud dari indikator.

Butir tes atau butir-butir pertanyaan harus sesuai dengan tujuan atau indikator, jelas, tidak ***ambiguous,*** singkat, dan menggunakan bahasa yang baku, bebas dari

bahasa lokal. Apabila ada gambar dan tabel maka gambar dan atau tabel itu harus jelas gambar dan fungsinya. Tidak perlu ada gambar dan atau tabel bila tidak diperlukan secara langsung. Untuk soal pilihan ganda maka pilihan jawaban yang disediakan harus homogen, baik panjang-pendeknya, jenis kata atau kalimatnya.

Dengan demikian setiap pernyataan atau butir soal perlu dibuat sedemikian rupa sehingga jelas apa yang ditanyakan dan jelas pula jawaban yang diminta. Mutu setiap butir soal akan menentukan mutu soal tes secara keseluruhan.

d. Telaah Soal atau Analisis Kualitatif Soal

Telaah soal atau analisis kualitatif soal adalah mengkaji secara teoritik soal tes yang telah tersusun. Telaah ini dilakukan dengan memperhatikan tiga aspek, yaitu aspek materi, aspek konstruksi, dan aspek bahasa.

e. Ujicoba Soal

Ujicoba soal pada dasarnya adalah upaya untuk mengetahui kualitas soal tes berdasarkan pada empirik atau respon dari peserta tes. Hal ini dapat terwujud manakala dilakukan analisis empirik atau analisis kuantitatif, baik menggunakan teori klasik maupun teori modern.

f. Perakitan Soal Tes

Agar skor tes yang diperoleh dapat dipercaya maka butir soal tes harus dibuat banyak. Oleh karena itu dalam penyajiannya butir-butir soal perlu dirakit menjadi alat ukur yang yang terpadu. Hal-hal yang dapat mempengaruhi validitas skor tes seperti urutan nomor soal, pengelompokan bentuk-bentuk soal, tata letak soal, dan sebagainya harus diperhatikan dalam perakitan soal menjadi sebiuah tes.

g. Penyajian Tes

Soal tes yang telah tersusun dapat langsung disajikan kepada peserta didik. Hal-hal yang perlu diperhatikan dalam penyajian tes ini adalah: waktu penyajian, petunjuk

yang jelas mengenai cara menjawab atau mengerjakan tes, ruangan dan tempat peserta didik. Pada prinsipnya, hal-hal yang menyangkut administratif penyajian tes harus diperhatikan agar pengetesan dapat berjalan lancar dan baik.

h. Skoring

Skoring atau pemeriksaan dan pemberian angka terhadap jawaban peserta didik merupakan langkah utama untuk mendapatkan informasi kuantitatif dari masing-masing siswa. Skoring harus objektif, dan tidak ada siswa yang dirugikan ataupun diuntungkan dengan cara yang digunakan. Apabila penyekoran dilakukan oleh dua orang yang tingkat kompetensinya sama, atau orang yang sama melakukan penyekoran dua kali maka hasilnya akan sama.

i. Pelaporan Hasil Tes

Pelaporan hasil penyekoran merupakan salah satu bentuk tanggung jawab pemberi tes kepada peserta didik, orang tua peserta didik, Keapala Sekolah, dan masyarakat. Hal ini sangat penting karena informasi ini dapat digunakan oleh sebagai bahan pertimbangan dalam menentukan pilihan, kebijakan, dan kebijaksanaan selanjutnya.

j. Pemanfaatan Hasil Tes

Informasi yang diperoleh dari pelaporan hasil penyekoran dapat dimanfaatkan sesuai dengan tujuan tes. Informasi ini dapat dimanfaatkan untuk perbaikan atau penyempurnaan sistem, proses atau kegiatan belajar mengajar maupun sebagai data untuk mengambil keputusan atau menentukan kebijakan.

3. Syarat Soal Tes yang baik

Seperti instrumen lainnya, soal tes juga harus baik, yakni memiliki validitas dan reliabilitas. Adapun penjelasan validitas dan reliabilitas adalah sebagai berikut.

a. Validitas

Validitas suatu alat ukur adalah sejauhmana alat ukur itu mampu mengukur apa yang seharusnya diukur (Nunnally, 1978). Sementara itu, Linn dan Gronlund (1995) menjelaskan validitas mengacu pada kecukupan dan kelayakan interpretasi yang dibuat dari penilaian, berkenaan dengan penggunaan khusus. Sedangkan Azwar (1996) menjelaskan suatu tes dapat dikatakan mempunyai validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang tepat dan akurat sesuai dengan maksud dikenakannya tes tersebut. Sisi lain yang sangat penting dalam konsep validitas adalah kecermatan pengukuran, yakni kemampuan untuk men-deteksi perbedaan-perbedaan kecil sekalipun yang ada pada atribut yang diukurnya.

Dalam pengukuran terhadap atribut psikologis, validitas sangat sulit dicapai. Hal ini dapat difahami karena pengukuran terhadap variabel psikologis dan sosial mengandung kesalahan yang lebih banyak daripada pengukuran variabel yang bersifat fisik. Oleh karena sulitnya menentukan validitas yang sebenarnya, maka yang dapat dilakukan adalah mengestimasi validitas instrumen dengan perhitungan tertentu.

Pengukuran psikologi itu mempunyai fungsi : (1) penegakan suatu hubungan statistik dengan variabel khusus, (2) representasi isi dari sesuatu, dan (3) pengukuran sifat-sifat psikologis. Oleh karenanya, validitas itu dapat dikelompokkan menjadi tiga tipe, yaitu: (1) validitas kriteria, (2) validitas isi, dan (3) validitas konstruk (Nunnally, 1978, Allen & Yen, 1979, Fernandes, 1984, Woolfolk & McCane, 1984, dan Lawrence, 1994).

Validitas berdasarkan kriteria dibedakan menjadi dua, yaitu validitas prediktif dan validitas konkuren. Fernandes (1984) mengatakan validitas berdasarkan kriteria

dimaksudkan untuk menjawab pertanyaan: *"How well test performance predicts future performance (predictive validity) or estimate current performance on some valued measure other than the test itself (concurrent validity)?"*. Hal senada juga disampaikan oleh Lawrence (1994) yang mengatakan bahwa tes dikatakan memiliki validitas prediktif bila tes itu mampu memprediksikan kemampuan yang akan datang. Dalam analisis validitas prediktif, performansi yang hendak diprediksikan disebut dengan kriteria. Besar kecilnya harga estimasi validitas prediktif suatu instrumen digambarkan dengan keofisien korelasi antara prediktor dengan kriteria tersebut.

Validitas isi suatu instrumen adalah sejauhmana butir-butir dalam instrumen itu mewakili komponen-komponen dalam keseluruhan kawasan isi objek yang hendak diukur dan sejauh mana butir-butir itu mencerminkan ciri perilaku yang hendak diukur (Fernandes, 1984; Nunnally, 1978). Sementara itu Lawrence (1994) menjelaskan bahwa validitas isi itu representativitas pertanyaan terhadap kemampuan khusus yang harus diukur.

Validitas konstruk adalah validitas yang menunjukkan sejauhmana instrumen mengungkap suatu trait atau konstruk teoretis yang hendak diukurnya (Fernandes, 1984; Nunnally, 1978). Prosedur validasi konstruk diawali dari suatu identifikasi dan batasan mengenai variabel yang hendak diukur dan dinyatakan dalam bentuk konstruk logis berdasarkan teori mengenai variabel tersebut. Dari teori ini ditarik suatu konskuensi praktis mengenai hasil pengukuran pada kondisi tertentu, dan konskuensi inilah yang akan dibuktikan secara empiris. Apabila hasilnya sesuai dengan harapan maka instrumen itu dianggap memiliki validitas konstruk yang baik.

Untuk tes hasil belajar, yang utama adalah validitas isi, yakni butir-butir soal yang ditanyakan kepada peserta didik sesuai dan mewakili kompetensi yang harus dicapai

oleh peserta didik. Hal ini dapat dilihat dari sejauh mana butir-butir soal itu sesuai dengan indikator yang merupakan jbaran dari kompetensi dasar.

b. Reliabilitas

Reliabilitas dapat diartikan sebagai keajegan atau kestabilan hasil pengukuran. Alat ukur yang reliabel adalah alat ukur yang mampu membuahkan hasil pengukuran yang stabil (Lawrence, 1994). Artinya suatu alat ukur dikatakan memiliki reliabilitas tinggi manakala digunakan untuk mengukur hal yang sama pada waktu berbeda hasilnya sama atau mendekati sama.

Reliabilitas alat ukur yang juga menunjukkan derajat kesalahan pengukuran tidak dapat ditentukan dengan pasti, melainkan hanya dapat diestimasi. Menurut Nunnally (1978) ada tiga cara mengestimasi reliabilitas, yaitu: (1) konsistensi internal, (2) tes paralel, dan (3) belah dua. Dalam cara konsistensi internal tes dilakukan hanya sekali pada sekelompok subjek kemudian dilakukan analisis atau diestimasi besarnya reliabilitas. Secara umum rumus untuk mengestimasi reliabilitas ini dapat digunakan rumus Koefisien Alpha. Namun apabila pilihan jawaban butir-butir pertanyaan/ pernyataan yang ada dalam instrumen/tes itu dikotomi maka dapat digunakan persamaan KR 20.

Tipe tes lainnya yang sering digunakan untuk mengestimasi reliabilitas adalah tipe tes paralel. Dalam tipe ini, tes dilakukan dua kali pada subjek yang sama namun tesnya berbeda meskipun paralel. Seperti yang telah dijelaskan di muka jarak antara ke dua tes ini sekitar dua minggu. Hasil kedua tes ini dikorelasikan, apabila koefisien korelasi ini kecil berarti tes itu kurang reliabel.

Selain konsistensi internal dan tes bentuk paralel, ada cara lain untuk mengestimasi reliabilitas, yaitu belah dua. Cara ini hanya menuntut satu kali tes untuk

subjek yang sama kemudian hasilnya dibelah dua. Idealnya pembelahan ini harus dilakukan secara random, namun adakalanya yang menggunakan cara skor dari butir-butir pertanyaan/pernyataan bernomor ganjil dipisahkan dengan skor dari butir-butir pertanyaan/pernyataan yang ber-nomor genap. Skor dari kelompok ini kemudian dikorelasikan dan selanjutnya digunakan rumus Spearman - Brown.

Salah satu cara untuk meningkatkan besarnya koefisien reliabilitas adalah memperpanjang tes, asalkan butir-butir yang ditambahkan harus homogen atau mengukur hal yang sama. Apabila butir yang ditambahkan tidak homogen maka reliabilitas tes tidak meningkat tetapi sebaliknya, malah menurun.

a. Analisis Soal Tes

Untuk mencapai butir-butir soal yang valid dan reliabel maka butir soal perlu dianalisis, yakni analisis secara teoritik atau telaah butir dan analisis kuantitatif untuk melihat tingkat kesulitan butir, daya beda butir, dan keberfungsian distraktor. Penjelasan analisis butir, baik kualitatif maupun kuantitatif adalah sebagai berikut.

1) Analisis Kualitatif (Telaah Butir)

Telaah kualitatif atau analisis teoritik dilakukan sebelum butir-butir soal diuji-cobakan dan di analisis secara empirik. Aspek-aspek yang diperhatikan dalam telaah kualitatif adalah aspek materi, konstruksi, dan bahasa/budaya ditelaah berdasarkan kaidah-kaidah yang telah ditentukan. Menurut Tim Pusbangsisjian, (1997/ 1998) kaidah-kaidah yang harus diperhatikan dalam menelaah butir soal yang berbentuk objektif pilihan ganda dapat dilihat pada Tabel 1 berikut.

Tabel 1. Lembar Telaah Butir Soal Pilihan Ganda

a)	**Aspek materi**
	(1) Soal sesuai dengan indikator; (2) Distraktor berfungsi; (3) Hanya ada satu kunci jawaban yang paling tepat
b)	**Aspek konstruksi**
	(1) Pokok soal dirumuskan dengan singkat, jelas dan tegas; (2) Rumusan pokok soal dan pilihan jawaban merupakan pertanyaan yang diperlukan (3) Pokok soal tidak memberi petunjuk ke kunci jawaban; (4) Pokok soal bebas dari pernyataan yang bersifat negatif ganda; (5) Gambar, grafik, tabel, diagram, wacana, dan sejenisnya yang terdapat pada soal jelas dan berfungsi; (6) Panjang pilihan jawaban relatif sama; (7) Pilihan jawaban tidak menggunakan pernyataan “Semua jawaban di atas salah” atau “Semua pilihan jawaban di atas benar” dan sejenisnya; (8) Pilihan jawaban yang berbentuk angka atau waktu harus disusun berdasar-kan urutan besar kecilnya angka tersebut atau kronologis; (9) Butir-butir soal tidak bergantung pada jawaban soal sebelumnya;
c)	**Aspek bahasa/budaya**
	(1) Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia; (2) Menggunakan bahasa yang komunikatif; (3) Tidak menggunakan bahasa yang berlaku setempat (bias budaya); (4) Pilihan jawaban tidak mengulang kata/kelompok kata yang sama.

Dalam analisis soal tes secara teoritik yang dikaji adalah kesesuaian antara butir-butir soal dengan tujuan atau indikator dan apakah soal tes sudah memenuhi validitas isinya. Soal tes juga dicermati penggunaan bahasa, kejelasan dan

kesingkatannya, juga dilihat kejelasan dan kefungsian tabel dan atau gambar. Pilihan jawaban juga dicermati homogenitas dan kejelasannya.

Selain kaidah untuk telaah butir secara teoritik, pedoman penyekoran juga harus jelas agar objektifitas pemberian skor oleh guru dapat dipertanggung-jawabkan. Pedoman pemberian skor untuk setiap butir soal uraian harus disusun sesegera mungkin setelah kalimat-kalimat butir soal tersebut selesai dirumuskan. Pedoman pemberian skor tidak boleh disusun saat koreksi akan dimulai.

Ada perbedaan pedoman penyekoran antara soal bentuk pilihan ganda dan soal bentuk uraian. Hal ini dikarenakan adanya perbedaan karakteristik di antara ke duanya yang secara rinci dapat dilihat pada Tabel 3 berikut.

Tabel 3. Perbandingan Antara Soal Bentuk Pilihan Ganda dan Uraian

Karakteristik	Uraian	Pilihan Ganda
Penulisan soal	Relatif mudah	Relatif sukar
Jumlah pokok bahasan yang ditanyakan	Terbatas	Lebih banyak
Aspek yang diukur	Dapat lebih dari satu	Hanya satu
Persiapan siswa	Penekanannya pada kedalaman materi	Lebih menekankan pada keluasan materi
Jawaban siswa	Mengorganisasikan jawaban	Memilih jawaban
Kecenderungan menebak	Tidak ada	Ada
Penyekoran	Sukar, lama, kurang konsisten (reliabel) dan subjektif	Mudah, cepat, sangat konsisten dan objektif

Pemilihan bentuk soal mana yang akan dipakai harus memperhatikan karakteristik soal seperti yang telah diuraikan di atas, tujuan penilaian dan efisiensi.

Untuk ujian yang jumlah pesertanya sangat banyak maka soal pilihan ganda lebih efisien, baik dilihat dari segi waktu maupun dari segi biaya yang dikeluarkan.

a. Analisis Kuantitatif

Analisis kuantitatif dilakukan dengan menggunakan dua pendekatan, yaitu pendekatan teori klasik dan pendekatan teori modern atau teori respon butir (Item Respon Theory =IRT). Dalam penlitian ini, hanya dijelaskan cara analisis butir kuantitatif dengan pendekatan teori klasik. Penjelasan analisis butir menurut teori klasik adalah sebagai berikut.

1) Pendekatan Teori Tes Klasik

Skor sebenarnya (*true score* = T) dan skor kesalahan (*error score* = E) adalah konstruk teoritik yang tidak dapat diamati. Hanya skor amatan (*observed score* = X) yang dapat diperoleh, dan skor amatan = skor sebenarnya + kesalahan (**X = T + E**). Jika kita berbicara skor sebenarnya, penting diingat bahwa skor sebenarnya yaitu skor rata-rata yang diperoleh dari pengulangan tes secara independen dengan menggunakan tes yang sama, adalah teoritis belaka. Skor ini tidak menunjukkan dengan lengkap karakteristik sebenarnya dari peserta tes kecuali kalau tes tersebut memiliki validitas sempurna, yaitu bahwa tes tersebut mengukur dengan tepat apa pokok isi yang diukur.

Menurut para ahli, ada beberapa kelemahan yang ada pada pendkatan teori klasik. Beberapa di antaranya adalah Hambleton, dkk (1991) dan Lord (1980). Mereka menjelaskan bahwa kelemahan-kelemahan tes teori klasik adalah: (1) statistik butir tes sangat tergantung pada karakteristik subjek yang dites; (2) taksiran kemampuan peserta tes sangat tergantung pada butir tes yang diujikan; (3) kesalahan baku penaksir skor berlaku untuk semua peserta tes, sehingga kesalahan baku pengukuran tiap peserta dan butir soal tidak ada; (4) informasi yang disajikan terbatas pada menjawab

benar atau salah saja tidak memperhatikan pola jawaban peserta tes; dan (5) asumsi tes paralel susah dipenuhi.

Walaupun teoriklasik ini memiliki beberapa kelemahan namun masih banyak yang menggunakan karena tidak menuntut responden besar (lebih 100) dan mudah mengaplikasikannya (melakukan analisis butir dengan pendekatan klasik ini). Oleh karenanya, untuk pengukuran yang melibatkan responden kecil misal pada pengukuran melalui tes harian pada bidang pendidikan, atau pengukuran pada bidang psikologi pada umumnya masih menggunakan pendekatan teori tes klasik.

Analisis kuantitatif menurut pendekatan teori tes klasik menghasilkan karakteristik butir yang meliputi tingkat kesukaran (p), daya pembeda (d), dan efektivitas distraktor. Selain itu, dengan analisis kuantitatif pendekatan teori klasik juga dapat diketahui reliabilitas soal tes, dan kesalahan baku pengukuran. Untuk melihat tingkat kesukaran, daya pembeda, dan efektivitas distraktor dilakukan analisis setiap butir tes, sedangkan reliabilitas dan kesalahan pengukuran baku dapat dilihat dengan cara menganalisis soal tes secara keseluruhan.

Kesesuaian karakteristik butir dengan jenis dan tujuan tes sangat menentukan kualitas butir tes. Pada analisis butir secara klasik, tingkat kesukaran (p) dapat diperoleh dengan beberapa cara, antara lain: (1) skala kesukaran linier; (2) skala bivariat; (3) indeks Davis; dan (4) proporsi menjawab benar. Cara yang paling mudah dan paling banyak digunakan adalah skala rata-rata atau proporsi menjawab benar atau *proportion correct* (p), yaitu jumlah peserta tes yang menjawab benar pada butir yang dianalisis dibandingkan dengan peserta tes seluruhnya.

Tingkat kesukaran *(p)* mengandung banyak kelemahan, antara lain tingkat kesukaran sebenarnya merupakan ukuran kemudahan butir karena semakin tinggi indeks *p*, semakin mudah butir tersebut. Sebaliknya semakin rendah *p* semakin sulit.

Oleh karenanya ada beberapa ahli pengukuran yang menyebut tingkat kesukaran ini dengan tingkat kemudahan. Tingkat kesukaran merupakan salah satu parameter butir soal, yang disimbolkan (P_i), yakni rasio antara jawaban benar dan banyaknya penjawab butir soal. Formulasi tingkat kesukaran butir soal adalah:

$$P_i = \frac{n}{N}$$

P_i = Tingkat kesukaran butir soal ke i
i = nomor butir soal
n = banyaknya siswa yang menjawab butir soal dengan benar
N = banyaknya siswa yang menjawab butir soal

Besarnya tingkat kesukaran berkisar antara nol dan satu. Suatu butir kadang-kadang dikategorikan ke dalam ekstrim sukar yaitu apabila nilai p mendekati nol dan ekstrim mudah apabila nilai p mendekati satu. Menurut Fernandes (1984), butir soal yang menghasilkan rerata skor sekitar 50 % dari skor maksimum dapat dikatakan bahwa butir soal itu mempunyai tingkat kesukaran yang tepat. Sementara itu, Thomas dan Dawson (1972) menjelaskan bahwa butir soal yang memiliki tingkat kesukaran 0,25 - 0,75 sudah dikatakan baik.

Daya pembeda atau daya beda suatu butir tes berfungsi untuk menentukan dapat tidaknya suatu butir tes membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Tujuan dari penelaahan daya pembeda adalah untuk melihat kemampuan butir tes tertentu dalam membedakan antara pengambil tes yang berkemampuan tinggi dan pengambil tes yang berkemampuan rendah.

Ada beberapa cara yang digunakan untuk menghitung daya pembeda, yaitu: (1) indeks diskriminasi, (2) indeks korelasi, dan (3) indeks keselarasan. Pada penelitian ini hanya dibahas dua cara untuk menghitung daya pembeda dengan metode korelasi

yaitu korelasi *point biserial* dan korelasi *biserial*. Korelasi *point biserial* maupun korelasi *biserial* adalah korelasi *product moment* yang diterapkan pada data, variabel-variabel yang dikorelasikan sifatnya masing-masing berbeda satu sama lain. Variabel butir tes bersifat dikotomi yaitu bernilai 1 untuk jawaban benar dan 0 jika jawaban salah. Di sisi lain, variabel skor total atau sub skor total bersifat kontinum yang diperoleh dari jumlah jawaban yang benar. Nilai koefisien korelasi *point biserial* selalu lebih jika dibandingkan dengan nilai koefisien korelasi *biserial*. Koefisien *point biserial* merupakan kombinasi hubungan antara butir tes, kriteria atau skor total, dan tingkat kesukaran. Korelasi *point biserial* cenderung lebih mengutamakan butir tes yang memiliki tingkat kesukaran rata-rata dan akan maksimum apabila tingkat kesukarannya *p* = 0.5 (Bahrul Hayat, 1996 dan Sumadi Suryabrata, 1987). Korelasi biserial merupakan korelasi antara butir tes dan kriteria, bebas dari pengaruh tingkat kesukaran butir tes. Menurut Crocker & Algina (1986) koefisien *point biserial* ditentukan dengan rumus:

$$\rho_{pbis} = \frac{\mu_+ - \mu_\tau}{\sigma_\tau} \sqrt{\frac{p}{q}}$$

ρ_{pbis} = Korelasi point biserial
μ_+ =Rerata-rata skor peserta tes yang menjawaban benar butir soal
μ_τ = Rerata skor total
σ_τ = Simpangan baku skor total
ρ = Proporsi banyaknya peserta yang menjawab benar
q = 1-ρ

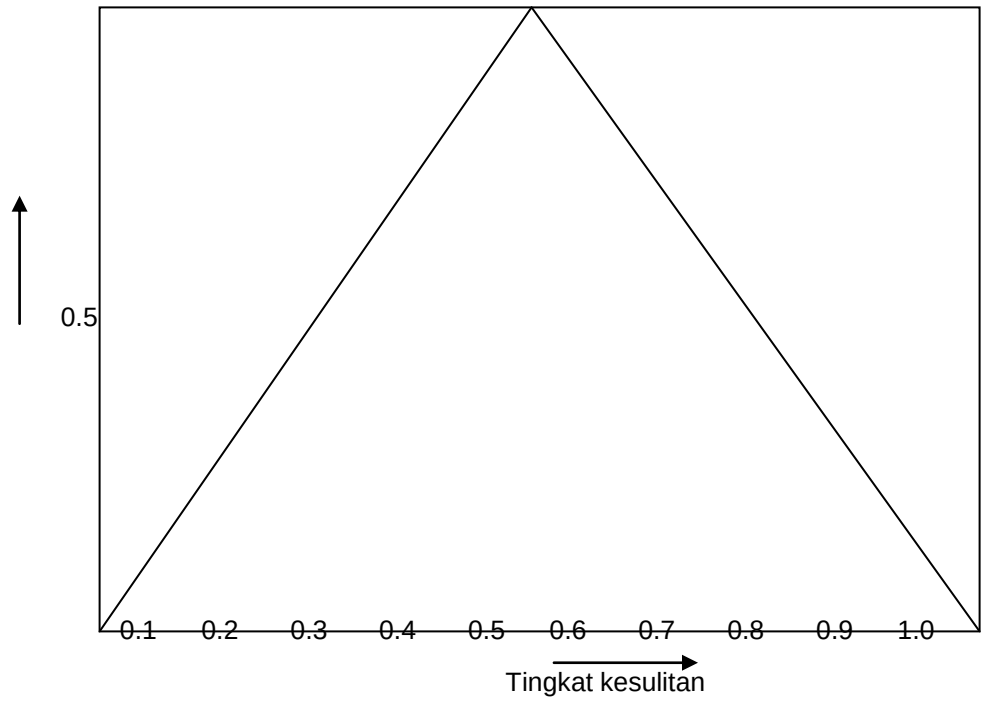
Sementara untuk menghitung indeks daya pembeda dengan korelasi *biserial* diguakan rumus::

$$\rho_{bis} = \frac{\mu_+ - \mu_\tau}{\sigma_\tau} \left(\frac{p}{Y}\right)$$

ρ_{bis} = Korelasi biserial
μ_+ = Rerata-rata skor peserta yang menjawab benar
μ_τ = Rerata skor total

σ_τ = Simpangan baku skor total
ρ = Proporsi banyaknya peserta yang menjawab benar
Y = Ordinat p dalam distribusi normal

Indeks daya beda butir soal dapat digunakan sebagai bahan pertimbangan sebuah butir baik atau tidak baik. Butir soal yang baik adalah butir soal yang mempunyai indeks daya beda lebih dari 0,2. seperti yang dinyatakan Fernandes(1984). Sementara Ebel (1972) menjelaskan suatu butir soal dikatakan berkualitas apabila indeks diskriminasi atau daya pembedanya paling sedikit 0,41. Selanjutnya, Fernandes (1984) menggambarkan hubungan antara tingkat kesukaran dan indeks daya pembeda seperti Gambar 1 berikut.



Gambar 1. Hubungan Antara Daya Pembeda dan Tingkat Kesulitan

Hal penting yang juga harus diperhatikan dalam menganalisis empirik butir soal adalah kemampuan distraktor atau alternatif jawaban yang disediakan menarik peserta tes untuk memilihnya. Jangan sampai tidak seorang peserta tes-pun memilih alternatif

jawaban yang disediakan. Fernandes (1984) yang mengutip pendapat Brawn menjelaskan distraktor dikatakan baik apabila paling tidak dipilih oleh 2 % dari seluruh peserta. Sementara itu, Nitko (1996) mengatakan distraktor dikatakan berfungsi manakala paling tidak dipilih oleh seorang peserta tes dari kelompok rendah. Pemilih dari kelompok rendah harus lebih banyak daripada kelompok atas. Distraktor juga dapat dikatakan berfungsi manakala peserta tes (siswa) dari kelompok atas dapat membedakan antara distraktor dan kunci jawaban sehingga yang memilih kunci jawaban lebih banyak daripada yang memilih distraktor.

Dalam menganalisis distribusi jawaban juga perlu memperhatikan kemungkinan salah kunci, yaitu manakala siswa dari kelompok atas yang memilih pengecoh lebih banyak daripada yang memilih kunci jawaban. Selain itu, juga perlu dideteksi ada tidaknya unsur tebakan dalam memilih alternatif jawaban. Hal ini dapat dilihat apabila jawaban peserta tes (siswa) merata, baik jawaban dari siswa kelompok atas maupun kelompok bawah.

Dengan demikian jelaslah bahwa untuk menilai kualitas butir tes tidak cukup hanya memperhatikan tingkat kesukaran dan daya pembeda butir tes yang bersangkutan. Penilaian kualitas butir tes juga harus melihat fungsi pilihan jawaban, terutama distraktor-distraktornya, yaitu harus tampak sebagai jawaban yang benar bagi subjek dari kelompok yang berkemampuan rendah. Sebaliknya harus tampak sebagai jawaban yang salah bagi subjek dari kelompok yang berkemampuan tinggi. Sekalipun suatu butir tes terlalu sukar atau terlalu mudah, namun apabila (1) daya pembeda butir tes, dan (2) distribusi jawaban, memenuhi kriteria, maka butir tes tersebut masih dapat diterima sebagai butir tes yang baik. Kriteria yang dimaksud adalah indeks daya pembeda butir tes $r_{bis} > 0,3$, dan indeks daya pembeda pilihan jawaban negatif kecuali kunci.

Hasil tes hendaknya juga membentuk distribusi normal. Hal ini dapat dicapai manakala butir-butir soal yang dipilih itu tepat, baik dilihat dari tingkat kesulitan maupun daya beda. Butir-butir soal yang tingkat kesulitannya tinggi cenderung menghasilkan skor yang memiliki distribusi juling positif atau hanya sebagian kecil peserta tes yang mendapat skor tinggi. Sebaliknya, bila butir-butir soal itu terlalu mudah maka skor yang diperoleh (hasil tes) cenderung membentuk juling negatif atau banyak sekali siswa yang mendapat skor tinggi.

DAFTAR PUSTAKA

Allen, M.J. & Yen, W.M. 1979. *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.

Croker, L. & Algina, J. 1986. *Introduction to classical and modern test theory*. New York : Holt, Rinehard and Winston Inc.

Dawson, J.B. & Thomas, G.H. 1972. *Item analysis and examination statics*. Birmingham: The Union of Educational Institutions.

Ebel, R.L. 1972. *Essentials of educational measurement*. (3rd. ed.) Englewood Cliffts,NJ: Prentice Hall Inc.

Fernandes, H.J. X. 1984. *Evaluation of educational program*. Jakarta: National Education Planning , Evaluating and Curriculum Development.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamentals of item response theory*. London: Sage Publications, Inc.

Bahrul Hayat. 2002. Sertifikasi, ujian akhir, dan otonomi daerah. *Makalah*. Disampaikan pada Seminar Sistem ujian akhir di era otonomi, di Jakarta, 19 Mei 2002.

Hayat, B. 1996. Interpretasi hasil analisis iteman. *Buletin Pengujian dan penilaian pendidikan*, 2 (5) 7 – 9.

Lawrence M.R. 1994. Question to ask when evaluaating test. *Eric digest*. Artikel: ED385607. Sumber: http://www.ericfacility.net/ericdigest/ ed.385607.html tanggal 10 Februari 2003.

Lord, F.M. 1980. *Applicatiom of item response theory to practical testing problems*. Hllsdale, NJ.: Lawrence Erlbaum Associates, Publisher.

Nitko, A.J. 1996. *Penilaian berkelanjutan berdasarkan kurikulum (PB2K): Kerangka, konsep, prosedur, dan kebijakan* (terj. AM. Ahmad) Jakarta: Pusat Pengembangan Agribisnis.

Nunnally, J.C. 1978. *Psychometric theory*. New York: McGraw Hill Book Company. Inc

Sumadi Suryabrata. 2000. *Pengembangan alat ukur psikologis*. Yogyakarta: Andi Offset.

Tim Sisjian. 1997/1978. *Bahan penataran Pengujian Pendidikan*. Jakarta: Pusbangsijian

Woolfolk, A.E. & McCune, L.N. 1984. *Educational Psychology for Teachers*. Englewood Cliffs, NJ: Prentice Hall, Inc.