

MENGANALISIS KUALITAS TES SEBAGAI SALAH SATU KOMPETENSI GURU PROFESIONAL

Oleh: H.Sujati¹

Abstrak

Mengevaluasi pengajaran merupakan salah satu kompetensi yang harus dikuasai oleh setiap guru yang professional. Selama ini kompetensi tersebut hanya dimaknai secara sempit, yakni menyusun, melaksanakan tes dan memberikan nilai berdasarkan skor hasil tes tersebut. Sementara kualitas tes itu sendiri hampir tidak pernah dipikirkan oleh guru. Tulisan ini berupaya memaparkan bagaimana suatu tes dianalisis, baik secara kualitatif maupun kuantitatif. Analisis tes secara kuantitatif meliputi penemuan indeks kesukaran butir, daya beda butir, keefektifan pengecoh (khusus untuk tes pilihan ganda) dan reliabilitas.

Kata Kunci: analisis hasil tes, kualitas tes, kompetensi guru

Pendahuluan

Pada akhir tahun 2005 ini menurut pengamatan penulis terdapat dua peristiwa penting yang terkait dengan dunia pendidikan, yakni disyahnkannya Undang-undang Guru dan ditingkatkannya anggaran untuk pendidikan. Pada tahun 2005 anggaran pendidikan nasional ditetapkan sebesar Rp.25,5 triliun dan pada tahun 2006

meningkat menjadi Rp 36 triliun. Dengan melihat besaran anggaran pendidikan nasional tersebut, pada saat ini Indonesia termasuk salah satu di antara sembilan negara tropis yang menaruh perhatian terhadap pendidikan. Pada tahun 2009 nanti diharapkan anggaran pendidikan nasional sudah mencapai 20% dari APBN (Kedaulatan Rakyat, 1 Desember 2005). Hal ini menunjukkan bahwa pemerintah memiliki keseriusan dalam menangani sektor pendidikan dan sekaligus menjadi indikator kesadaran bahwa pendidikan merupakan investasi kemanusiaan jangka panjang yang sangat menentukan kemajuan bangsa.

Disyahnkannya Undang-undang Guru dan kenaikan anggaran pendidikan pendidikan sebenarnya merupakan variabel tak langsung untuk menuju pendidikan yang bermutu. Di antara kedua variabel tersebut masih banyak sekian banyak variabel antara yang perlu diperhatikan. Kompetensi guru misalnya, merupakan variabel tergantung yang diharapkan kena dampak langsung dari dua perubahan tersebut. Disyahnkannya Undang-undang Guru dan kenaikan anggaran pendidikan menjadi tidak bermakna apabila tidak meningkatkan etos kerja dan kinerja guru dalam menjalankan tugas profesinya.

¹ H. Sujati adalah Dosen FIP UNY

Pada garis besarnya setiap guru memiliki tiga tugas utama, yaitu merencanakan, melaksanakan dan mengevaluasi pembelajaran (Cece Rakhmat dan Didi Suherdi, 1999). Sementara itu Depdiknas (2002) merekomendasikan empat kompetensi yang seharusnya dimiliki oleh seorang guru, yakni menguasai bidang studi, memahami peserta didik, menguasai pembelajaran yang mendidik, dan mampu mengembangkan kepribadian dan keprofesionalan. Dalam konteks ini, kemampuan melaksanakan evaluasi termasuk dalam kompetensi pembelajaran yang mendidik. Di sini kelihatan bahwa kemampuan melaksanakan evaluasi merupakan salah satu dimensi dari kompetensi guru.

Selama ini terdapat pandangan yang kurang tepat tentang evaluasi pengajaran. Evaluasi pengajaran dimaknai secara sempit, yakni menyusun tes, melaksanakannya dan memberikan nilai berdasarkan skor hasil tes tersebut. Di sini terdapat satu langkah penting yang sering dilupakan guru, yakni menganalisis kualitas tes. Padahal dengan tidak diketahuinya kualitas tes tersebut dapat menyebabkan keputusan yang diambil oleh guru menjadi bias atau sesat. Di sinilah letak pentingnya seorang guru memiliki kompetensi menganalisis kualitas tes.

Kualitas Tes

Menurut Depdikbud (1993) tes yang baik harus memenuhi dua kriteria, yakni kriteria kuantitatif dan kualitatif. Yang termasuk kriteria kualitatif adalah bahasa, konstruksi dan materi. Sementara kriteria kuantitatif adalah bersifat baku, memiliki daya pembeda yang baik, memiliki tingkat

kesukaran yang baik, dan memiliki reliabilitas baik. Kualitas suatu tes ditentukan oleh kualitas butir-butir soal yang membangun tes itu.

Kaplan dan Saccuzo (dalam Depdikbud, 1993) menyatakan bahwa untuk menguji kualitas butir soal perlu dilakukan analisis butir soal. Tujuan analisis butir soal adalah mendapatkan informasi tentang baik-buruknya suatu butir soal berdasarkan kriteria-kriteria tertentu dan pada akhirnya mengetahui kualitas tes secara keseluruhan.

Menurut Depdikbud (1993) analisis butir soal dapat dilakukan secara kualitatif dan kuantitatif. Analisis secara kualitatif pada dasarnya merupakan penelaahan soal yang secara teoritis dikaji dari sudut pandang isi atau materi uji dan teknis penulisan soal. Oleh karena itu analisis kualitatif sering juga disebut analisis validitas logis (Sumarna Supranata, 2004). Kajian teoritis materi uji meliputi kesesuaian materi dengan jenjang dan tingkat sekolah siswa yang hendak diuji dan kesesuaian materi uji dengan indikator yang tertuang dalam kisi-kisi tes. Kajian teknis penulisan butir soal memfokuskan tampilan soal ditinjau dari sudut pandang kaidah penulisan soal. Disamping itu, kaidah pemakaian bahasa Indonesia juga perlu diperhatikan.

Analisis butir soal secara kuantitatif menekankan pada analisis karakteristik internal tes melalui data yang diperoleh secara empiris. Analisis tersebut meliputi penentuan tingkat kesukaran, daya pembeda, distribusi jawaban dan reliabilitas (Depdikbud, 1993). Analisis butir secara kuantitatif khusus untuk soal bentuk pilihan ganda selain diarahkan pada tingkat kesukaran daya pembeda dan validitas, juga per

memperhatikan keefektifan pengecoh (*distractor*) (Sumarna Supranata, 2004). Oleh karena itu pada pembahasan lebih lanjut akan diuraikan mengenai tingkat kesukaran, daya pembeda, keefektifan pengecoh dan reliabilitas.

1. Tingkat kesukaran butir

Menurut Oller (dalam Burhan Nurgiyantoro (2001) tingkat kesukaran butir (*item difficulty*) merupakan pernyataan tentang seberapa sulit atau seberapa mudah soal bagi peserta yang dikenai tes. Dalam hal ini Oller justru menggunakan istilah tingkat kemudahan (*item facility*) karena yang terjadi sebenarnya bukan menggambarkan seberapa sulit, tetapi justru sebaliknya, yakni seberapa mudah.

Butir soal yang baik adalah butir yang tidak terlalu sulit dan tidak terlalu mudah. Butir yang sangat sulit atau mudah sama tidak baiknya, karena keduanya tidak mampu membedakan antara siswa yang berkemampuan tinggi dan berkemampuan rendah (Noll dalam Burhan Nurgiyantoro, 2001).

Menurut teori klasik, tingkat kesukaran merupakan proporsi antara peserta tes yang menjawab butir tersebut secara benar dan banyaknya peserta tes secara keseluruhan (Sumarna Supranata, 2004). Tingkat kesulitan suatu butir dinyatakan dengan suatu indeks yang berkisar antara 0,0 sampai 1,0, Indeks 0,0 menunjukkan bahwa butir itu terlalu sulit, tidak ada seorangpun yang mampu menjawabnya secara benar. Sebaliknya, indeks 1,0 menunjukkan bahwa butir tersebut sangat mudah sehingga semua peserta tes mampu menjawab benar butir tersebut.

Seberapa besar indeks kesukaran yang ideal itu? Saifuddin Azwar (1996) memberikan suatu patokan bahwa indeks kesukaran yang ideal adalah yang berada pada sekitar 0,50. Sementara itu Oller (dalam Burhan Nurgiyantoro, 2001) menyatakan bahwa suatu butir dikatakan layak apabila memiliki indeks kesukaran berkisar 0,15 sampai 0,85. Indeks yang berada di luar itu dinyatakan sebagai butir yang tidak layak, sehingga perlu direvisi atau digugurkan. Suharsimi Arikunto (1991) membuat klasifikasi indeks kesukaran menjadi tiga jenjang, yakni: sukar ($\leq 0,29$), sedang (0,30 – 0,70) dan mudah (≥ 71)

Indeks kesukaran tes bentuk pilihan ganda dapat dicari dengan rumus:

$$IK = \frac{B}{JS}$$

Keterangan:

IK adalah indeks kesukaran yang dicari

B adalah jumlah siswa yang menjawab butir itu secara benar

JS adalah jumlah siswa peserta tes

Sementara itu indeks kesukaran tes esai menurut Burhan Nurgiyantoro (2001) dapat dicari dengan rumus:

$$IK = \frac{\sum Bt + \sum Br - (2N \times Skor_{min})}{2N \times (Skor_{maks} - Skor_{min})}$$

Keterangan:

IK = indeks kesukaran yang dicari

$\sum Bt$ = skor betul kelompok tinggi

$\sum Br$ = jumlah skor betul kelompok rendah

Skor_{maks} = skor maksimal suatu butir soal
Skor_{min} = skor minimal suatu butir
N = jumlah kelompok tinggi atau rendah

Formula di atas hanya berlaku untuk mencari indeks kesukaran tes bentuk pilihan ganda, sedangkan untuk tes bentuk esai harus dicari dengan formula yang berbeda. Burhan Nurgiyantoro (2001:147) memberikan rumus sebagai berikut:

Indeks kesukaran =

$$\frac{\sum Bt + \sum Br - (2N \times Skor_{min})}{2N \times (Skor_{maks} - Skor_{min})}$$

Keterangan:

$\sum Bt$ = Jumlah skor betul kelompok tinggi
 $\sum Br$ = Jumlah skor betul kelompok rendah
Skor_{maks} = Skor maksimal suatu butir soal
Skor_{min} = Skor minimal suatu butir
N = Jumlah subyek kelompok tinggi atau rendah

2. Daya pembeda

Secara umum daya pembeda diartikan sebagai kemampuan suatu butir untuk membedakan antara peserta tes yang berkemampuan tinggi dan berkemampuan rendah (Burhan Nurgiyantoro, 2001). Suatu butir dikatakan baik apabila butir tersebut dapat dijawab benar oleh sebagian besar peserta tes yang berkemampuan tinggi dan hanya dapat

dijawab benar oleh sebagian kecil dari peserta tes yang berkemampuan rendah. Butir tes yang dapat dijawab benar atau salah oleh peserta tes yang berkemampuan tinggi dan berkemampuan rendah menunjukkan bahwa tes tersebut tidak memiliki daya beda.

Daya pembeda dihitung berdasarkan perbedaan jumlah jawaban benar untuk setiap butir soal antara peserta tes yang berkemampuan tinggi dan berkemampuan rendah. Jika terjadi kelompok peserta tes yang berkemampuan rendah lebih banyak menjawab benar dari pada peserta tes yang berkemampuan tinggi, menunjukkan bahwa butir soal tersebut perlu direvisi atau diganti. Sebaliknya, apabila peserta kelompok tinggi lebih banyak menjawab benar, hal itu menunjukkan bahwa butir itu baik.

Besarnya indeks daya beda dinyatakan dengan suatu angka yang besarnya berkisar antara -1,0 sampai dengan 1,0. Indeks daya beda -1,0 berarti seluruh peserta tes kelompok bawah menjawab benar butir tersebut, sementara seluruh peserta tes kelompok atas menjawab salah. Indeks daya beda 1,0 menunjukkan bahwa butir tersebut dijawab benar oleh seluruh peserta tes kelompok atas, sementara seluruh peserta tes kelompok bawah menjawab salah.

Menurut Oller (dalam Burhan Nurgiyantoro, 2001) butir soal yang baik adalah yang setidaknya memiliki indeks daya pembeda 0,25. Indeks daya pembeda di bawah angka tersebut menunjukkan bahwa butir tersebut tidak layak, sehingga perlu direvisi atau digugurkan.

Indeks daya beda tes bentuk pilihan ganda dapat dicari dengan rumus:

$$DB = \frac{B_a}{J_a} - \frac{B_b}{J_b}$$

Keterangan:

DB = daya beda yang dicari

B_a = jumlah peserta kelompok atas yang menjawab benar

B_b = jumlah peserta kelompok bawah yang menjawab benar

J_a = jumlah kelompok atas

J_b = jumlah kelompok bawah

Satu formula lain untuk mencari daya beda adalah melalui pendekatan biserial dan point biserial (Saifuddin Azwar, 1996). Namun karena pendekatan biserial lebih kompleks maka yang diperkenalkan dalam tulisan ini hanya pendekatan point biserial. Suharsimi Arikunto (1991) menyebut pendekatan tersebut sebagai upaya mencari validitas butir. Rumus point biserial adalah:

$$\gamma_{pbi} = \frac{Mp - Mt}{St} \sqrt{\frac{p}{q}}$$

Keterangan:

γ_{pbi} = koefisien validitas yang dicari

Mp = Rerata skor dari subyek yang menjawab betul pada item yang dicari validitasnya

Mt = rerata skor total

S = Simpangan baku

p = Proporsi siswa yang menjawab butir itu benar

q = proporsi siswa yang menjawab butir itu salah

Berbagai formula di atas hanya cocok untuk mencari daya beda pada tes berbentuk obyektif yang berskala dikotomi, sementara untuk tes esai yang berskala kontinum menggunakan formula lain. Menurut Burhan Nurgiyantoro (2001:147) rumus untuk mencari daya beda tes esai adalah:

$$DB = \frac{\sum Bt - \sum Br}{N(Skor_{maks} - Skor_{min})}$$

Keterangan:

DB = Daya Beda yang dicari

$\sum Bt$ = jumlah skor betul kelompok tinggi

$\sum Br$ = jumlah skor betul kelompok rendah

$Skor_{maks}$ = Skor maksimal suatu butir soal

$Skor_{min}$ = Skor minimal suatu butir

N = Jumlah subyek kelompok tinggi atau rendah

3. Analisis keefektifan pengecoh

Analisis pengecoh bertujuan mengetahui tingkat keberfungsian pengecoh yang disediakan. Dengan mengetahui penyebaran jawaban dapat diketahui: (a) pengecoh yang terlalu menyolok kesalahannya sehingga tidak ada yang memilih, (b) pengecoh yang menyesatkan, yakni pengecoh yang lebih banyak dipilih oleh siswa kelompok atas dari pada siswa kelompok bawah, dan (c) pengecoh yang memiliki daya tarik bagi siswa kelompok bawah. Dengan demikian dapat dinyatakan bahwa pengecoh sebenarnya berfungsi untuk membedakan antara

siswa yang berkemampuan tinggi dan rendah. Pengecoh dikatakan berfungsi efektif apabila banyak dipilih oleh siswa yang berkemampuan rendah

Baik-buruknya butir soal tidak saja dilihat dari sudut tingkat kesukaran dan daya beda butir. Dalam banyak kasus, ketidakbaikan butir soal justru dipengaruhi oleh ketidakefektifan pengecoh yang disajikan. Pengecoh yang disajikan ternyata tidak memiliki daya pikat sehingga tidak ada satupun peserta tes yang terjebak. Untuk itu perlu dilakukan analisis penyebaran frekuensi jawaban pada alternatif jawaban yang disediakan. Apabila dari 40 peserta tes misalnya, tidak ada satupun yang memilih alternatif jawaban tertentu yang berperan sebagai pengecoh, hal ini menunjukkan bahwa alternatif tersebut tidak efektif, sehingga perlu diganti atau direvisi. Suharsimi Arikunto (1991) memberi rambu bahwa pengecoh dikatakan efektif apabila mampu menjebak sekurang-kurangnya 5% dari seluruh peserta tes. Dalam program Iteman keefektifan pengecoh ini disebut Proporsi Endorsing (PE) yang dicari dengan rumus:

$$PE = \frac{p_i}{JS} \times 100\%$$

Keterangan:

p_i = jumlah pemilih pada pengecoh tertentu

JS = Jumlah peserta tes

4. Reliabilitas

Kriteria lain dari tes yang baik adalah memiliki tingkat kepercayaan atau reliabilitas secara memadai. Reliabilitas berkaitan dengan ketetapan hasil pengukuran. Suatu tes dinyatakan reliabel apabila tes itu dikenakan pada subyek yang sama dalam kurun waktu yang berbeda memberikan hasil (skor) yang kurang lebih sama. Allen dan Yen (1979) menyatakan bahwa reliabilitas merupakan kestabilan skor yang diperoleh orang yang sama ketika diuji ulang dengan menggunakan tes yang sama pada situasi yang berbeda. Tuckman (dalam Burhan Nurgiyantoro, 2001) menyatakan bahwa kepercayaan tes menunjuk pada pengertian apakah suatu tes dapat mengukur secara konsisten sesuatu yang diukur dari waktu ke waktu. Selanjutnya dinyatakan bahwa konsistensi itu terkait dengan hal-hal sebagai berikut: (1) tes dapat memberikan hasil yang relatif tetap terhadap sesuatu yang diukur, (2) jawaban siswa terhadap butir-butir tes secara relatif tetap, dan (3) tes tersebut diperiksa oleh siapapun akan memberikan hasil yang kurang lebih sama.

Menurut Depdikbud (1993) terdapat dua macam konsistensi reliabilitas tes, yakni konsistensi internal dan eksternal. Konsistensi internal merupakan suatu kriteria di mana ketetapan ditentukan berdasarkan hasil tes itu sendiri. Sementara itu, pada konsistensi eksternal, ketetapan ditentukan dengan cara mengkorelasikan dengan hasil tes yang

lain. Suatu tes dinyatakan reliabel apabila hasil tes pertama dan kedua memiliki korelasi tinggi. Sebaliknya, apabila tes pertama dan kedua korelasinya rendah maka dinyatakan tes itu tidak reliabel.

Adanya dua pandangan konsistensi tersebut menyebabkan terjadinya berbagai jenis reliabilitas tes. Fernandes (dalam Burhan Nurgiyantoro, 2001) menyebutkan tiga jenis reliabilitas, yakni: (1) konsistensi internal, yang terdiri dari empat macam: Alpha, Kuder Richardson 20, Kuder Richardson 21 dan belah dua Spearman-Brown; (2) stabilitas yang dilakukan dengan cara uji ulang; dan (3) equivalensi yang dilakukan dengan cara membuat tes paralel. Dari ketiga jenis reliabilitas tersebut dalam tulisan ini hanya jenis yang pertama, yakni konsistensi internal yang dibahas. Dalam konteks karya tulis ini terdapat tiga teknik untuk mencari indeks reliabilitas, yakni: Kuder Richardson (KR)₂₀, Kuder Richardson (KR)₂₁, dan Alpha. KR₂₀ dan KR₂₁ digunakan untuk mencari reliabilitas tes bentuk obyektif, sedangkan Alpha biasanya digunakan untuk tes bentuk esai. Di bawah ini ditulis ketiga formula reliabilitas tersebut.

$$KR_{20} = \left(\frac{k}{k-1} \right) \left(\frac{S^2 - \sum pq}{S^2} \right)$$

Keterangan:

KR₂₀ = indeks reliabilitas yang dicari

k = banyaknya butir

S = simpangan baku

p = proporsi jawaban benar

q = 1-p

$$KR_{21} = \left(\frac{k}{k-1} \right) \left(1 - \frac{M(k-M)}{kS_i^2} \right)$$

Keterangan:

KR₂₁ = indeks reliabilitas yang dicari

k = banyak butir

M = rerata skor total

S_i² = varian total

$$\text{Alpha} = \frac{k}{k-1} \left(1 - \frac{\sum S_i^2}{S^2} \right)$$

Keterangan:

k = jumlah butir soal

$\sum S_i^2$ = jumlah varian butir-butir soal

S_i² = varian total

Besarnya koefisien reliabilitas berkisar antara -1 sampai dengan 1. Koefisien reliabilitas 0 atau bahkan negatif menunjukkan bahwa tingkat kepercayaan tes tersebut rendah. Sebaliknya, koefisien reliabilitas tinggi menunjukkan bahwa tingkat kepercayaan tes itu juga tinggi. Menurut Tuckman, tes buatan guru dinyatakan reliabel apabila memiliki koefisien $\geq 0,60$, sedang untuk tes yang dipublikasikan paling tidak 0,85 (Burhan Nurgiyantoro, 2001).

Tinggi rendahnya koefisien reliabilitas dipengaruhi oleh berbagai faktor. Menurut Crokder dan Algina (dalam Depdikbud, 1993) koefisien reliabilitas dipengaruhi oleh panjang

tes, tingkat kesukaran, homogenitas kelompok dan daya beda butir. Namun menurut penelitian Aiken (dalam Depdikbud, 1993) yang paling banyak mempengaruhi koefisien reliabilitas adalah tingkat kesukaran. Hal ini karena menyangkut variasi jumlah soal yang dijawab benar. Sementara menurut Saifuddin Azwar (1996) tinggi-rendahnya koefisien reliabilitas dipengaruhi oleh panjang tes dan daya beda butir. Tes yang dibangun oleh banyak butir yang berdaya beda tinggi cenderung memiliki tingkat reliabilitas tinggi. Sebaliknya, konstruksi tes yang dibangun oleh butir-butir soal yang berdaya beda rendah cenderung memiliki tingkat reliabilitas rendah atau bahkan negatif.

Penutup

Untuk mengakiri tulisan ini, penulis perlu mengingatkan sekali lagi bahwa profesionalitas seorang guru salah satunya ditentukan oleh kompetensinya dalam melaksanakan penilaian pengajaran. Penilaian pengajaran cakupannya tidak saja meliputi menyusun, melaksanakan tes dan memberikan nilai berdasarkan skor hasil tes. Satu langkah kegiatan yang tidak kalah penting adalah melakukan analisis terhadap kualitas tes. Analisis tersebut dapat bersifat kualitatif dan kuantitatif. Analisis kualitatif berupaya melihat kualitas tes dari sudut kaidah penyusunan tes dan bahasa yang digunakan. Sementara itu analisis kuantitatif meliputi: menemukan indeks kesukaran butir, daya beda, efektivitas pengecoh (khusus untuk tes pilihan ganda) dan reliabilitas.

Daftar Pustaka

- Allen, M.J. dan Yen, Wend M. (1979). *Introduction to Measurement Theory*. California: Brooks Publishing Company.
- Burden, Paul R. dan Byrd, David M. (1999). *Methods for Effective Teaching*. Boston: Allyn and Bacon.
- Cece Rachmadi dan Didi Suherdi. (1999). *Evaluasi Pengajaran*. Jakarta: Depdikbud.
- Depdikbud. (1993). *Analisis Soal Secara Klasik*. Jakarta: Dirjen Dikdasmen.
- Depdiknas. (2002). *Standar Kompetensi Guru Kelas SD-MI*. Jakarta: Dirjen Dikti.
- Perrott, Elizabeth. (1985). *Effective teaching: A practical guide to improving your teaching*. New York: Longman.
- Saifuddin Azwar. (1996). *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.
- Suharsimi Arikunto. (1991). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Sumarna Supranata. (2004). *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes*. Bandung: Rosda.

