

STRATEGI PENYUSUNAN INSTRUMEN TES KELOMPOK MATAKULIAH KETEKNIKAN

Sumarno*)

Abstrak

Untuk mengetahui apakah seorang mahasiswa telah memiliki kompetensi berkenaan dengan materi yang telah diajarkan diperlukan adanya tes. Untuk mengetahui hal ini seorang dosen dituntut untuk menyusun atau mengembangkan tes yang berkualitas. Berkenaan dengan mata kuliah keteknikan tes yang dapat dikembangkan dapat berupa tes perbuatan (menekankan aspek psikomotor) maupun tes kognitif. Instrumen tes yang memenuhi kriteria daya pembeda, reliabilitas, validitas dan tingkat kesukaran butir akan menghasilkan data sesuai dengan konstruk kompetensi yang menjadi fokus dan yang akan diukur oleh tes tersebut. Tulisan ini dimaksudkan sebagai guide bagi mahasiswa, dosen atau peneliti dalam mengembangkan instrument tes sehingga diperoleh instrument non yang memenuhi syarat untuk mengumpulkan data.

Kata kunci: Penyusunan instrument tes

Pendahuluan

Strategi penyusunan instrument tes diartikan sebagai seperangkat kebijaksanaan/ketentuan yang dipilih, yang telah dikaitkan dengan faktor yang menentukan warna dari instrument tersebut, yang mencakup pemilihan bentuk instrumen dan penyajian instrument yang memenuhi kriteria sebagai instrument yang memenuhi syarat sebagai alat pengumpul data atau mengetahui kompetensi seseorang. Berkenaan dengan hal tersebut, pada tulisan ini disajikan syarat-syarat instrument tes yang baik, langkah-langkah pengembangan, dan kaidah penulisan instrumen tes.

A. Tes dan Syarat-Syarat Tes Yang Baik

Secara singkat Allen dan Yen (1979) menyatakan tes adalah alat atau *device* untuk mendapatkan sampel kemampuan seseorang. Menurut Depdiknas (1999) tes adalah himpunan pertanyaan yang harus dijawab atau pertanyaan-pertanyaan yang harus dipilih/ditanggapi, atau tugas-tugas yang harus dilakukan oleh orang yang dites (*testee*) dengan tujuan untuk mengukur suatu aspek (perilaku) tertentu dari orang yang dites. Dalam tes prestasi belajar yang

hendak diukur adalah tingkat kemampuan peserta tes dalam menguasai bahan pelajaran yang telah disampaikan/diajarkan kepadanya.

Tes untuk mengukur prestasi, baik untuk aspek pengetahuan maupun aspek keterampilan, dapat dibagi menjadi dua macam, yaitu tes kemampuan (*power test*) dan tes kecepatan (*speed test*).

a. Prinsip dari tes kemampuan adalah tidak adanya batasan waktu di dalam mengerjakan tes. Menurut prinsip ini, jika waktu pengerjaan tes tidak dibatasi, maka hasil tes benar-benar mengungkap kemampuan seseorang. Adanya pembatasan waktu dalam mengerjakan tes, diperkirakan akan menyebabkan orang menjadi tidak dapat menunjukkan seluruh kemampuan yang dimilikinya, sehingga skor yang dihasilkan tidak menunjukkan seluruh kemampuan yang sebenarnya dari orang tersebut. Namun demikian tidak efisien jika peserta yang paling lambat pun harus ditunggu dalam menyelesaikan butir tes. Sebagai kompromi, para ahli tes menyediakan

*)Prof. Dr.Sumarno, M.Pd, adalah Dosen Jurusan Teknik Mesin FT Unimed

waktu yang diperkirakan sebagian besar dapat menyelesaikan tes tersebut.

- b. Sebaliknya pada tes kecepatan, yang diukur adalah kecepatan di dalam memikirkan atau mengerjakan butir tes. Butir tes itu biasanya relatif mudah, sehingga yang diukur benar-benar kecepatan bekerja atau berpikir seseorang bukan kemampuan lainnya.

Pada bagian di atas telah dinyatakan bahwa tes adalah alat untuk mengukur kemampuan seseorang. Seperti halnya alat ukur panjang (misalnya mistar), alat ukur berat (timbangan), dan sebagainya. Sebagai alat ukur aspek kemampuan manusia, tes juga mempunyai syarat sebagai alat ukur pada umumnya, yaitu:

a. Validitas (*Unidimensi*)

Tes yang telah dikembangkan hanya mengukur satu dimensi atau satu aspek saja. Jadi sebuah tes hasil belajar, misalnya Ilmu Bahan, maka setiap soalnya harus mengukur hanya pengetahuan Ilmu Bahan saja. Pada alat ukur fisika, memenuhi syarat uni dimensi sangat mudah, tetapi pada butir tes Ilmu Bahan tidak mustahil terdapat unsur bahasa, atau bentuk butir tes yang melibatkan istilah yang belum dikenal oleh peserta tes. Sehingga berhasil tidaknya peserta dalam menjawab butir tersebut masih akan tergantung pada aspek lain selain pengetahuannya tentang Ilmu Bahan. Untuk mengetahui syarat ini dapat dilakukan dengan mencermati masing-masing butir, adakah kata-kata baru yang belum pernah dipahami oleh sebagian besar peserta tes. Dari sudut pengukuran syarat ini terpenuhi apabila setiap butir berkorelasi satu sama lainnya, sehingga dapat dijadikan bukti bahwa semua aspek tersebut memang

merupakan bagian dari suatu yang lebih umum atau luas.

Menurut Carmines dan Zeller (1979) validitas menggambarkan sejauhmana sebuah instrumen pengukuran mengukur apa yang ingin diukur. Seseorang memvalidasi berarti bukan mengukur instrumen itu sendiri tetapi mengukur instrumen dalam hubungannya dengan untuk apa instrumen tersebut digunakan. Menurut Azwar (1992) instrumen yang menghasilkan data yang relevan dengan tujuan pengukuran dikatakan sebagai instrumen yang memiliki validitas. Terkandung disini pengertian bahwa valid tidaknya instrumen tergantung pada mampu tidaknya instrumen tersebut mencapai tujuan pengukuran yang dihendaki dengan tepat. Oleh karena itu, validitas tidak berlaku untuk umum untuk semua tujuan pengukuran. Suatu instrumen biasanya hanya mengukur satu tujuan spesifik. Dengan demikian predikat valid suatu instrumen harus diikuti oleh keterangan yang menunjuk pada kepada tujuan ukur, yaitu valid untuk mengukur apa. Misalnya instrumen ini valid untuk mengukur kompetensi ilmu bahan.

Pengertian lain validitas adalah kecermatan pengukuran. Suatu instrumen yang valid tidak sekedar mampu mengungkap data dengan tepat, tetapi juga harus memberikan gambaran yang cermat tentang data tersebut. Cermat berarti pengukuran itu mampu memberi gambaran mengenai perbedaan yang sekecil-kecilnya diantara subjek yang satu dengan lainnya (Azwar, 1992).

Berdasarkan cara estimasinya dan disesuaikan dengan sifat dan fungsi setiap instrumen, tipe validitas dapat dikelompokkan menjadi tiga kelompok, yaitu: *content validity* (validitas isi), *construct validity* (validitas konstruk), dan

criterion-related validity (validitas berdasarkan kriteria). Validitas isi terbagi menjadi dua tipe, yaitu: *face validity* (validitas wajah), dan *logical validity* (validitas logik). Validitas wajah adalah validitas yang paling rendah signifikasinya karena hanya didasarkan pada penilaian terhadap format penampilan. Validitas logik mengungkap suatu *trait* atau kontrak teoritik yang hendak diukur (Allen & Yen, 1979). Untuk menguji validitas konstruk diawali dengan batasan mengenai ubahan yang hendak diukur, kemudian ubahan itu dinyatakan sebagai suatu konstruk logis menurut konsep-konsep teoritis dari ubahan tersebut. Dari teori ini kemudian ditarik semacam konsekuensi praktis mengenai hasil (data) pada kondisi tertentu. Apabila setelah diuji hasilnya sesuai dengan konsep-konsep teoritis, maka instrumen tersebut dianggap memiliki validitas konstruk yang baik.

Validitas berdasar kriteria menghendaki tersedianya kriteria eksternal yang dapat dijadikan dasar pengujian instrumen tersebut. Prosedur validasi berdasarkan kriteria menghasilkan dua macam validitas, yaitu validitas prediktif (*predictive validity*) dan validitas konkuren (*concurrent validity*). Validitas prediktif sangat penting artinya bila instrumen dimaksudkan untuk memprediksi performansi di waktu yang akan datang. Validitas konkuren dilakukan dengan mensejajarkan instrumen yang kita buat dengan instrumen lain yang telah terlebih dahulu teruji validitasnya.

b. Keandalan (reliabilitas)

Keandalan tes mencakup ketepatan/kecermatan (*precision*) hasil pengukuran, dan keajegan/kestabilan (*consistency*). Sebuah tes dengan jumlah

disebut juga dengan validitas sampling. Validitas tipe ini menunjuk pada sejauhmana isi instrumen merupakan representasi dari cirri-ciri atribut yang hendak diukur.

Validitas konstruk adalah tipe validitas yang menunjukkan sejauhmana tes butir yang banyak dan seluruh butir tes memiliki tingkat kesukaran yang “sedang” (*on target*) bagi peserta, tentu akan menghasilkan informasi yang lebih teliti jika dibandingkan dengan tes yang jumlah butirnya sedikit dan tingkat kesukarannya *off target*. Dengan kata lain, butir tes tidak boleh terlalu jauh di bawah atau di atas kemampuan tingkat pencapaian belajar peserta tes, dan tingkat kesukaran butir tes sebaiknya homogen.

Keajegan dapat dibedakan menjadi dua, yaitu keajegan internal dan keajegan eksternal. Keajegan internal adalah tingkat sejauhmana butir-butir tes tersebut homogen baik dari segi tingkat kesukaran maupun dari segi bentuk soal dan prosedur menjawabnya. Keajegan eksternal adalah tingkat sejauhmana skor yang dihasilkan dari penyajian sebuah tes kepada sekelompok orang akan tetap sama sepanjang kemampuan peserta yang diukur tersebut sama atau belum berubah.

Tujuan menghitung reliabilitas instrumen adalah untuk mengetahui tingkat ketepatan (*precision*) atau keajegan (*consistency*) skor pengukuran. Keandalan suatu instrumen memberikan informasi tentang besarnya kesalahan pengukuran. Keandalan suatu instrumen dapat dikategorikan menjadi tiga, yaitu: konsistensi internal, stabilitas, dan antar penilai. Besarnya indeks konsistensi internal diperoleh dari data hasil uji coba. Untuk mencari indeks ini cukup dilakukan satu kali uji coba. Indeks stabilitas merupakan tingkat kestabilan hasil

pengukuran yang dilakukan paling tidak dua kali untuk orang yang sama dalam waktu yang berbeda. Keandalan antar penilai diperoleh dari besarnya korelasi hasil penskoran dari dua orang terhadap lembar jawaban yang sama. Besarnya indeks keandalan ini adalah 0 sampai 1, dan yang dapat diterima minimum 0,7. Makin tinggi koefisien reliabilitas suatu instrumen (mendekati 1), maka makin tinggi keajekannya (Depdiknas, 2002).

c. Memiliki Daya Deskriminan

Butir tes kompetensi keteknikan yang bermutu baik akan membantu dosen dalam meningkatkan pelaksanaan proses belajar mengajar. Butir yang bermutu baik dapat memberikan informasi yang tepat tentang mahasiswa mana yang belum atau sudah menguasai materi yang telah diajarkan. Karakter butir yang dapat membedakan kelompok siswa yang telah menguasai materi yang telah diajarkan dengan yang belum menguasai disebut dengan daya deskriminan (daya beda). Semakin tinggi kemampuan mahasiswa dalam memahami materi yang telah diajarkan, maka semakin tinggi peluang menjawab dengan benar. Semakin rendah kemampuan mahasiswa dalam memahami materi yang telah diajarkan, maka semakin kecil peluang menjawab dengan benar.

D. Langkah-Langkah Penyusunan Tes

Test yang telah didesign dapat digunakan sebagai alat seleksi/penerimaan, klasifikasi/*placement*, evaluasi/diagnosis, dan konseling. Tes dapat digunakan untuk seleksi ketika ingin dilakukan pemilihan/pemisahan sekelompok orang dari kelompok orang tertentu. Misalnya tes penerimaan mahasiswa baru. Tes dapat digunakan untuk klasifikasi/*placement*, adalah tes yang digunakan untuk

mengelompok orang ke dalam beberapa kelompok. Tes dapat digunakan untuk evaluasi/diagnosis ketika peserta didik ditetapkan kemampuannya di dalam kelas. Tes dapat digunakan dalam konseling, ketika kita gunakan inventori minat untuk memberi saran berkenaan dengan karir-karir potensial atau menggunakan tes personalitas untuk membantu dalam konseling pernikahan (Allen dan Yen, 1979).

Menurut Allen dan Yen (1979) menyebutkan ada delapan langkah dasar dalam pengembangan tes, dan menurut Nur (1987) dan Depdiknas (1999) ada 10 langkah. Pada bagian ini akan disajikan hasil pencermatan terhadap ketiga sumber tersebut sehingga diharapkan diperoleh perangkat tes yang memadai untuk mengukur kompetensi keteknikan. Langkah-langkah penyusunan tes secara garis besar dikelompokkan menjadi empat yaitu perencanaan, penulisan butir, penelaahan butir tes (kualitatif dan kuantitatif), dan perakitan butir tes.

1. Perencanaan tes

a. Menentukan tujuan tes

Seperti dipaparkan di atas tujuan tes bermacam-macam. Masih diragukan bahwa sebuah tes tunggal dapat dikembangkan untuk memenuhi semua tujuan tes secara maksimal. Tes yang dikembangkan untuk seleksi cenderung memiliki tingkat kesukaran yang tinggi. Tes klasifikasi, digunakan untuk membedakan peserta tes di sepanjang rentang kemampuan yang lebar harus terdiri atas butir tes yang memiliki tingkat kesukaran sedang. Tes diagnostik digunakan untuk menentukan daerah kelambanan tertentu dari peserta tes harus terdiri atas butir tes dengan tingkat kesukaran mudah.

b. Mempersiapkan area/materi yang akandicover

Aktivitas ini merupakan hal yang sangat penting untuk mendapatkan validitas logis dan validitas sampling. Validitas logis dan validitas sampling adalah versi validitas tampilan (*face validity*) yang njlimet (mutahir). Hal ini mencakup definisi domain dari tingkah laku/kompetensi yang akan diukur melalui tes dan design logis

dari butir untuk mengcover semua area penting dari domain tersebut. Validitas logis khususnya berguna di dalam mengembangkan tes prestasi.

Pada langkah ini kembangkan kisi-kisi domain untuk tes. Kisi-kisi materi ajar juga merupakan deskripsi mengenai ruang lingkup dan isi dari apa yang akan diujikan, yang terdiri atas matrik pokok bahasan dan sub pokok bahasan (Tabel 1).

Tabel 1. Kisi-Kisi Materi Ajar

SUB POKOK BAHASAN	POKOK BAHASA			
	1	2	3	4
I				
II				

Setiap sel akan mewakili area konten dari sel. Tes akan dikembangkan dengan proporsi yang sama atau tidak, tergantung pada evaluasi pengembang tes terhadap tingkat kepentingan relatif dari sel tersebut. Penentuan materi yang akan diujikan merupakan bagian yang sangat penting karena di dalam suatu ujian tidak mungkin semua materi/indikator yang telah diajarkan dapat diujikan dalam waktu yang terbatas, misalnya dalam waktu satu atau dua jam. Oleh karena itu, setiap dosen harus menentukan materi mana yang sangat penting dan penunjang, sehingga dalam waktu yang sangat terbatas, materi/indikator yang diujikan hanya menanyakan materi-materi yang sangat penting saja. Penentuan **materi penting** dilakukan dengan memperhatikan kriteria:

1. Urgensi, yaitu materi secara teoritis mutlak harus dikuasai oleh mahasiswa.
2. Kontinuitas, yaitu materi lanjutan yang merupakan pendalaman dari satu atau lebih materi yang sudah dipelajari sebelumnya.
3. Relevansi, yaitu materi yang diperlukan untuk mempelajari atau memahami mata kuliah lainnya.
4. Keterpakaian, yaitu materi yang memiliki nilai terapan tinggi dalam kehidupan sehari-hari.

c. Menentukan kompetensi yang akan diujikan

Menentukan standar kompetensi merupakan acuan/target utama yang harus dipenuhi melalui setiap kompetensi dasar yang ada. Dalam menentukan kompetensi yang akan diukur, penulis dapat menentukan atau memperhatikan jenis kata kerja yang telah dikembangkan oleh Bloom, atau tokoh lain seperti: Quellmalz, Mazano, Gagne, Krathwohl, Gronlund dan Maclay, atau Linn dan Gronlund.

Untuk kata kerja ranah kognitif yang dikembangkan oleh Bloom adalah: (1) level ingatan, diantaranya: menyebutkan, menentukan, menunjukkan, mengingat kembali, mendefinisikan. (2) level pemaham, diantaranya: membedakan, mengubah, memberi contoh, memperkirakan, mengambil kesimpulan. (3) level penerapan, diantaranya: menggunakan, menerapkan, (4) level analisis, diantaranya: membandingkan, mengklasifikasikan, mengkategorikan, menganalisis, (5) level sintesis, diantaranya: menghubungkan, mengembangkan, mengorganisasikan, menyusun, dan (6) level evaluasi, diantaranya: menafsirkan, menilai, dan memutuskan.

d. Menetapkan penyebaran butir tes berdasarkan: kompetensi dan materi

Setelah kegiatan penentuan materi yang akan ditanyakan selesai, kegiatan berikutnya adalah menentukan secara tepat penyebaran butir tes berdasarkan: kompetensi dan materi. Kompetensi yang akan diukur tergantung pada tuntutan kompetensi, baik standar kompetensi maupun kompetensi dasarnya. Setiap kompetensi di dalam kurikulum/silabus memiliki tingkat keluasan dan kedalaman kemampuan yang tidak sama. Makin tinggi kemampuan yang diukur, maka makin sulit soalnya. Perilaku yang akan diukur dapat dilihat pada "*perilaku yang terdapat pada rumusan kompetensi dasar atau pada standar kompetensi*". Di samping itu dapat pula dilihat pada "*perilaku yang terdapat pada indikator pencapaian hasil belajar/hasil belajar*" yang tertera pada kurikulum/silabus. Bila ingin mengukur perilaku yang lebih tinggi, dosen dapat mendaftar terlebih dahulu semua perilaku yang paling sulit/tinggi, berdasarkan rumusan kompetensinya (baik kompetensi dasar maupun kompetensi standar). Dari

susunan perilaku itu, dapat dipilih satu perilaku yang tepat diujikan kepada mahasiswa.

e. Menyusun Kisi-Kisi Butir Tes

Kisi-kisi ini dikenal dengan istilah “*test blue-print*” atau “*table of specification*”. Kisi-kisi tes adalah format berbentuk matrik yang membuat informasi yang dijadikan pedoman untuk menulis atau merakit butir tes menjadi perangkat tes. Tujuan penyusunan kisi-kisi adalah untuk menentukan ruang lingkup dan tekanan tes yang setepat-tepatnya, sehingga menjadi petunjuk dalam menulis soal. Wujud dan bentuk format atau matriks kisi-kisi ditunjukkan sebagai berikut (Tabel 2).

No	Kompetensi dasar	Hasil belajar/ Indikator	Materi Pokok	No. soal
1	2	3	4	5

Tabel 2
FORMAT KISI-KISI PENULISAN BUTIR TES

Keterangan:

Isi pada kolom 2, 3, dan 4 harus sesuai dengan pernyataan yang ada di dalam kurikulum/silabus.

Kisi-kisi yang baik harus sesuai dengan persyaratan sebagai berikut:

1. Kisi-kisi harus dapat mewakili isi silabus/kurikulum atau materi yang telah diajarkan secara tepat dan proporsional.
2. Komponen-komponennya diuraikan secara jelas dan mudah dipahami.
3. Materi yang hendak ditanyakan dapat dibuatkan soalnya.

Indikator dalam kisi-kisi merupakan pedoman dalam merumuskan soal yang dikehendaki. Indikator yang baik adalah indikator yang dirumuskan secara singkat dan jelas. Syarat indikator yang baik adalah:

1. Menggunakan kata kerja operasional (perilaku khusus) yang tepat.
2. Menggunakan satu kata kerja operasional untuk soal objektif, dan lebih dari satu kata kerja operasional untuk soal uraian atau perbuatan.
3. Dapat dibuatkan soal atau pengecohnya (untuk soal objektif).

2. Menulis Butir Tes

Penulisan butir soal untuk tes tertulis merupakan suatu kegiatan yang sangat penting dalam menyiapkan bahan ujian. Setiap butir soal yang ditulis harus berdasarkan rumusan indikator soal yang sudah disusun di dalam kisi-kisi dan berdasarkan kaidah penulisan soal.

Menulis soal bentuk pilihan ganda sangat diperlukan keterampilan dan ketelitian. Hal yang paling sulit dilakukan dalam menulis soal bentuk pilihan ganda adalah menuliskan pengecohnya. Pengecoh yang baik adalah pengecoh yang tingkat kerumitan atau tingkat kesederhaan, serta panjang pendeknya relatif sama dengan kunci jawaban. Oleh karena itu, untuk memudahkan dalam penulisan soal bentuk pilihan ganda, maka perlu mengikuti langkah-langkah sebagai berikut: pertama, menuliskan pokok soalnya, langkah kedua menuliskan kunci jawabannya, dan kemudian menuliskan pengecohnya.

Pedoman umum dalam penulisan butir soal pilihan ganda adalah: (1) pokok soal harus jelas, (2) pilihan harus homogen, (3) panjang kalimat pilihan jawaban diusahakan sama, (4) tidak ada petunjuk yang mengarahkan ke pilihan jawaban yang benar, (5) menghindari pemakaian pilihan jawaban seperti: semua benar, ketiganya salah, atau sejenisnya, (6) semua pilihan jawaban memiliki hubungan yang logis dengan pokok soal, (7) tidak menggunakan kata

negatif ganda, (8) memakai bahasa baku, dan (9) letak jawaban benar ditentukan secara acak.

3. Menelaah Butir Secara Kualitatif

Validasi butir tes secara kualitatif dilakukan sebelum butir tes diujikan/digunakan. Aspek yang perlu diperhatikan dalam memvalidasi butir tes secara kualitatif adalah berkenaan dengan aspek materi, konstruksi, bahasa, dan kunci jawaban (Lampiran 1). Ada dua cara/teknik yang dapat ditempuh untuk menelaah (memvalidasi) butir tes secara kualitatif, yaitu: teknik moderator dan teknik panel.

Teknik moderator merupakan teknik berdiskusi yang di dalamnya terdapat satu orang penengah/wasit. Berdasarkan teknik ini, setiap butir tes didiskusikan secara bersama-sama dengan beberapa ahli, seperti dosen yang mengajar materi, ahli materi, penyusun/pengembang kurikulum, ahli tes, ahli bahasa, dan berlatar belakang psikologi. Para penelaah dipersilahkan mengomentari/memperbaiki berdasarkan ilmu yang dimilikinya. Setiap komentar/masukan dari peserta diskusi dicatat oleh notulis. Setiap butir dituntaskan secara bersama-sama. Kelemahan teknik ini adalah memerlukan waktu yang lama untuk mendiskusikan setiap satu butir tes.

Teknik panel merupakan suatu teknik dimana setiap butir tes ditelaah berdasarkan kaidah-kaidah penulisan butir tesnya, yaitu ditelaah dari segi materi, konstruksi, bahasa/budaya, kebenaran kunci jawaban/pendoman penskoran yang dilakukan oleh beberapa penelaah (Lampiran 1).

Tabel 3

Resume Hasil Penilaian Butir Tes

No. Butir	Skor Penilai					Mean	Ket
	A	B	C	D	E		
1							Baik
2							Diperbaiki
3							Diganti
Dst							

Caranya adalah beberapa penelaah diberi butir-butir tes yang akan ditelaah, format penelaahan, dan pedoman penilaian/penelaahannya. Para penelaah dipersilakan memperbaiki langsung pada teks butir tes dan memberi komentar serta memberi nilai pada setiap butir tes yang kriterianya adalah: baik, diperbaiki, diganti. Hasil resume dari para penelaah disajikan seperti ditunjukkan pada Tabel 3.

4. Merakit butir-butir tes menjadi perangkat tes

Merakit soal adalah menyusun soal yang siap pakai menjadi satu perangkat/paket. Untuk memudahkan pelaksanaannya, penulis soal harus memperhatikan langkah-langkah sebagai berikut:

1. Mengelompokkan soal yang mengukur kompetensi dan materi yang sama, kemudian soal-soal itu ditempatkan dalam urutan yang sama.
2. Memberi nomor urut soal didasarkan nomor urut soal dalam kisi-kisi
3. Mengecek setiap soal dalam satu paket apakah soal-soal sudah mengikuti kaidah.
4. Membuat petunjuk umum dan khusus untuk mengerjakan soal.
5. Membuat format jawaban, lembaran kunci dan petunjuk penilaian

5. Menelaah Butir Secara Kuantitatif

Penelaahan butir secara kuantitatif adalah penelaahan butir berdasarkan pada data empirik dari butir tes yang bersangkutan. Penelaahan butir tes secara kuantitatif dilakukan setelah tes diujikan. Data diperoleh dari peserta yang mengerjakan butir tes tersebut. Ada dua pendekatan atau model teori yang

dapat digunakan untuk menganalisis butir tes secara kuantitatif, yaitu: pendekatan secara klasik atau tradisional, dan teori respon modern atau *Item Response Theory* (Izard, 1995).

Aspek yang diperhatikan dalam analisis butir soal klasik adalah setiap butir tes ditelaah dari segi: tingkat kesukaran soal, daya pembeda, penyebaran pilihan jawaban (untuk soal bentuk objektif) atau frekuensi jawaban pada setiap pilihan jawaban. Ciri klasik ditunjukkan oleh kenyataan bahwa kelompok butir pada tes tidak dapat dipisahkan dari kelompok peserta yang menempuh ujites (Naga, 1992).

a. Tingkat Kesukaran Butir

Secara matematis dapat dinyatakan bahwa tingkat kesukaran butir soal adalah proporsi antara jawaban benar dari peserta dengan keseluruhan peserta. Indeks kesukaran butir pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0,00 sampai 1,00 (Aiken, 1994). Makin besar indeks tingkat kesukaran (TK) yang diperoleh dari hasil suatu perhitungan, berarti makin mudah soal itu. Suatu butir tes memiliki TK = 0,00 artinya bahwa tidak ada peserta yang menjawab dengan benar butir tes tersebut, dan bila TK = 1 artinya bahwa semua peserta menjawab dengan benar. Menurut Nitko (1996: 310) rumus TK dinyatakan sebagai berikut:

$$TK = \frac{\text{Jumlah peserta yang menjawab benar suatu butir tes}}{\text{Jumlah peserta yang mengikuti tes}}$$

Hasil perhitungan dengan menggunakan rumus di atas menggambarkan TK dari butir tes yang tersebut. Sebagai pedoman umum,

klasifikasi TK tes dapat dikelompokkan sebagai berikut:

TK 0,00 – 0,30 butir terestergolong sukar.

TK 0,31 – 0,70 butir tes tergolong sedang.

TK 0,71 – 1,00 butir terestergolong mudah.

b. Daya Pembeda

Daya pembeda butir tes adalah kemampuan suatu butir tes dapat membedakan antara peserta tes yang telah menguasai materi dan peserta yang belum atau tidak menguasai materi. Manfaat yang dapat diambil dari diketahuinya daya pembeda adalah:

1. Untuk meningkatkan mutu setiap butir tes melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir tes dapat diketahui apakah butir tes itu baik, direvisi, atau ditolak.
2. Untuk mengetahui seberapa jauh setiap butir dapat mendeteksi/membedakan kemampuan peserta, yaitu peserta yang telah memahami atau belum memahami materi yang telah disampaikan oleh guru/dosen. Menurut Safari (2004) apabila butir soal tidak dapat membedakan kedua kemampuan peserta tersebut, maka butir tes dapat dicurigai “kemungkinannya” sebagai berikut:
 - a. Kunci jawaban butir tes itu tidak tepat.
 - b. Butir tes memiliki dua atau lebih jawaban yang benar.
 - c. Kompetensi yang diukur tidak jelas.
 - d. Pengecoh tidak berfungsi.
 - e. Materi yang ditanyakan terlalu sulit, sehingga banyak peserta yang menebak.

Indeks daya pembeda setiap butir tes dinyatakan dalam bentuk proporsi. Makin tinggi indeks daya pembeda butir tes, makin mampu butir tes tersebut membedakan antara peserta yang telah memahami materi dengan peserta yang belum memahami materi yang telah diajarkan. Indeks pembeda berkisar antara $-1,00$ sampai dengan $+1,00$. Makin tinggi indeks pembeda tersebut makin baik butir tes tersebut. Jika daya pembeda < 0 berarti lebih banyak kelompok bawah (peserta yang kurang memahami materi) menjawab butir tes tersebut dengan benar dibandingkan dengan kelompok atas (peserta yang memahami materi). Secara matematis daya pembeda dapat dirumuskan sebagai berikut:

$$DP = \frac{2(BA - BB)}{N}$$

Dengan

DP : daya pembeda butir tes.

BA : jumlah jawaban benar dari kelompok atas

BB : jumlah jawaban benar dari kelompok bawah

N : Jumlah peserta yang mengikuti tes.

Menurut Crocker dan Algina (1986: 315), klasifikasi indeks daya pembeda dapat dikelompokkan menjadi empat, yaitu:

- 0,40 – 1,00 butir tes diterima/baik
- 0,30 – 0,39 butir diterima tetapi perlu diperbaiki
- 0,20 – 0,29 butir diperbaiki
- 0,00 – 0,19 butir tidak dipakai/dibuang.

c. Penyebaran Jawaban

Penyebaran pilihan jawaban dijadikan dasar dalam penelaahan butir. Hal ini dimaksudkan untuk mengetahui berfungsi tidaknya jawaban yang tersedia. Suatu pilihan jawaban

(pengecoh) dapat dikatakan berfungsi apabila pengecoh:

1. Paling tidak dipilih oleh 5% peserta tes/siswa.
2. Lebih banyak dipilih oleh kelompok siswa yang belum paham materi.

d. Reliabilitas Instrumen Tes (pilihan ganda)

Untuk mengetahui koefisien reliabilitas tes bentuk pilihan ganda dipergunakan rumus Kuder Richardson 20 (KR-20), sebagai berikut:

$$KR - 20 = \frac{k}{k - 1} \frac{\sum p(1 - p)}{(SD_x)^2}$$

Ket: k = jumlah butir tes

p = proporsi jawaban benar

$(1 - p)$ = proporsi jawaban salah

e. Validitas Intrumen Tes (pilihan ganda)

Untuk mengetahui validitas butir bentuk pilihan ganda digunakan persamaan korelasi point biserial, sebagai berikut:

$$rpbis = \frac{xb - xs}{SDt} \frac{pq}{\bar{p}}$$

Ket: xb = rerata skor peserta yang menjawab benar

xs = rerata skor peserta yang menjawab salah

SDt = simpangan baku skor total

6. Perakitan butir berdasarkan hasil analisis

Butir yang telah memenuhi syarat yang mencakup reliabilitas, validitas, dan daya beda dirakit menjadi satu perangkat tes, sedangkan butir yang tidak memenuhi

syarat perlu pertimbangan apakah perlu diperbaiki atau dibuang.

E. Kesimpulan

Agar diperoleh instrumen tes yang memenuhi syarat diperlukan adanya kompetensi dari pengembang tes tentang bagaimana syarat-syarat tes yang baik dan langkah-langkah yang harus ditempuh sehingga diperoleh instrumen tes yang baik. Tes yang baik minimal harus memiliki validitas, reliabilitas, dan daya deskriminan (daya beda).

Langkah yang harus seorang pengembang instrumen tes mencakup: perencanaan, penulisan butir tes, analisis kualitatif, perakitan butir tes, analisis kuantitatif, dan perakitan final.

DAFTAR PUSTAKA

- Aiken, Lewis R. (1994). *Psychological testing and assessment*. 8th Edition. Boston: Allyn and Bacon.
- Allen, M. Y dan Yen, W. M. (1979). *Intruduction of measurement theory*. Berkeley, California: Brooks/Cole Publishing Company.
- Azwar, Saifuddin. (1992). *Reliabilitas dan validitas*. Yogyakarta : Pustaka Pelajar.
- Carminer, E. G dan Zeller, R.A (1979). *Reliability and validity assessment*. Beverly Hill: Sage Publications.
- Crocker, L dan Algina, J. (1986). *Introduction to clasical and modern test theory*. New York: Holt, Rinehart and Winston, Inc.
- Depdiknas. (2002). *Pengembangan instrumen dan penilaian ranah efektif*. Jakarta: Direktorat Pendidikan Menengah Umum. Depdiknas.
- Izard, John. (1995). *Trial testing and item analysis* (Module C4). Australia: Australian Council for Educational Research. Unesco.
- Nitko, Anthony J. (1996). *Educational assessment of students*. 2nd Edition. Ohio: Merril Publishing Company.
- Nur, Muhammad. (1987). *Pengantar teori tes*. Jakarta: Dirjen Dikti. Departemen Pendidikan dan Kebudayaan.
- Safari. (2004). *Teknik analisis butir soal instrumen tes dan non tes*. Jakarta: Direktorat Pendidikan Menengah Umum.