

TAHAPAN PENYUSUNAN TES

Membuat sebuah tes yang baru memerlukan ilmu dan seni. Pengembang tes harus memilih strategi dan materi, dan kemudian melakukan penelitian yang terus menerus yang akan mempengaruhi kualitas instrumen yang akan dihasilkan. Dalam penyusunan tes secara garis besar ada enam langkah yang harus ditempuh untuk sampai kepada menghasilkan sebuah tes. Langkah-langkah tersebut adalah mendefinisikan tes, memilih metode skala, menyusun item, menguji item, merevisi tes, dan menerbitkan tes.

Mendefinisikan tes

Mendefinisikan tes adalah membatasi skop dan tujuannya, serta tahu perbedaan alat ukur yang akan dibuat dengan alat ukur-alat ukur psikologis yang ada sebelumnya. Akan sia-sia jika kita membuat alat ukur yang skop dan tujuannya sama dengan tes sebelumnya dan alasan apa yang membuatnya menyusun tes tersebut.

Salah satu contoh tes yang menyatakan bahwa tes yang disusun oleh pengembangnya itu berbeda dengan tes-tes yang ada sebelumnya adalah Kaufman Assessment Battery for Children (K-ABC). Penyusun tes mencantumkan enam alasan mengapa atau tujuan dia menyusun tes inteligensi ini, yaitu:

1. Mengukur inteligensi dengan dasar teoritis dan penelitian yang kuat
2. Memisahkan pengetahuan faktual dari kemampuan memecahkan masalah yang tidak biasa
3. Memperoleh skor dari intervensi pendidikan
4. Memasukkan tugas-tugas baru]
5. Mudah diadministrasikan dan skornya obyektif
6. Bisa diterapkan pada anak semua anak prasekolah, kelompok minoritas, dan anak berkebutuhan khusus.

Dari enam penjelasan tadi dapat diperoleh gambaran bahwa Kaufman memiliki tujuan yang jelas dan hal yang baru dalam mengukur inteligensi jauh sebelum dia menyusun tes itu.

Memilih Metode Skala

Ada banyak skala psikologi yang dikembangkan selama ini. Tujuan tes psikologi adalah penugasan angka-angka kepada jawaban-jawaban tes sehingga jawaban itu bisa dinilai baik atau buruk dari apa yang diukur. Aturan penandaan angka itu disebut dengan metode skala. Pemilihan skala tergantung kepada kecocokan skala itu dengan apa yang diukur, jadi tidak ada istilah “bahwa skala yang satu lebih baik dari skala yang lain”.

Dalam skala ini memerlukan pemahaman yang baik bagi penyusun tes tentang level pengukuran. Seperti yang telah kita bahas di bab sebelumnya bahwa dalam pengukuran ada empat level pengukuran, yaitu nominal, ordinal, interval dan rasio. Dengan paham level pengukuran ini akan mendukung penyusun dalam melakukan analisis statistik. Misalnya, ketika dia melakukan korelasi dua variabel, dia bisa memilih teknik korelasi apa yang harus dipilih yang sesuai dengan level pengukuran dari variabel yang diukur. Berikut adalah uraian sekilas dari metode skala yang pernah dikembangkan oleh ahli pengukuran psikologi.

1. Expert ranking

Metode skala ini digunakan ketika tidak mungkin dilakukan pemberian pertanyaan kepada subyek yang diukur. Misalnya mengukur kedalaman koma seseorang yang mengalami di kepala. Untuk mengetahui sejauhmana kedalaman koma pasien maka diperlukan seorang ahli dalam bidang ini atau dokter untuk menilai kedalaman koma pasien.

2. Method of Equal-Appearing Interval

Metode ini diciptakan oleh Thurstone yang disebut sebagai bapak skala psikologi. Metode ini membutuhkan analisis statistik yang agak rumit namun secara logis bisa dengan dipahami. Ada lima langkah untuk menyusun skala ini, yaitu

- Mengumpulkan sebanyak mungkin pernyataan Ya-Tidak yang mengungkapkan suatu sikap yang variatif yaitu positif sampai yang negatif kepada suatu fenomena atau obyek.
- Kumpulkan 10 orang ahli atau penilai untuk menilai pernyataan-pernyataan ini untuk menentukan tingkat favorabilitas (dukungan) atau unfavorabilitas (penolakan) dalam bersikap. Para penilai itu diminta memberi nilai dengan rentang 1 sampai 11. semakin besar nilai yang dia beri maka penilaiannya terhadap pernyataan itu semakin favorabel. Para penilai ini tidak diminta untuk merespons sesuai apa yang rasakan terhadap sikap yang diukur tapi diminta menilai kefavorabelan sikap yang dinyatakan dalam pernyataan-pernyataan itu.
- Setelah setiap pernyataan dinilai, carilah median dan interkuartil (Q3-Q1) dari setiap nilai pernyataan itu. Item yang dipilih adalah item dengan interkuartil terkecil. Median dari item yang terpilih adalah bobot atau skor dari item tersebut.
- Setelah item-item terpilih (11 item terpilih) diberi tanda pilihan Ya dan Tidak. Responden diminta memilih jawaban Ya atau Tidak.
- Setiap item yang dijawab Ya oleh responden maka akan diberi skor sesuai dengan bobotnya dan kemudian dijumlahkan semua bobot item yang dijawab Ya oleh responden.

3. Likert Scale

Skala Likert menyodorkan kepada responden lima atau tujuh jawaban yang diurutkan mulai dari tidak setuju sampai setuju.

Saya tahu mengapa saya sedih	SS	S	R	TS	STS
Saya tidak tahu apa yang harus saya lakukan ketika ada teman yang mengeluh tentang permasalahannya	SS	S	R	TS	STS

	Skor Favorable	Skor Unfavorable
SS : Sangat Setuju	4	0
S : Setuju	3	1
R : Ragu-Ragu	2	2
TS : Tidak Setuju	1	3
STS: Sangat Tidak Setuju	0	4

Skor akhir dari skala ini adalah skor total dari semua item yang ada dalam skala tersebut. Skala ini disebut juga dengan skala mengukur perbedaan individu.

4. Guttman Scale

Skala Guttman, mengasumsikan bahwa responden yang memilih satu pernyataan maka akan dipastikan memilih pula pernyataan yang tingkatannya lebih rendah pada kontinum yang sama.

Contoh:

- () Saya kadang-kadang merasa sedih
- () Saya sering merasa sedih
- () Saya hampir selalu merasa sedih
- () Saya selalu merasa sedih

Jika responden memilih pernyataan: Saya hampir selalu merasa sedih, maka dia juga akan setuju dengan pernyataan dengan tingkat yang lebih rendah lainnya yaitu pernyataan: Saya kadang-kadang merasa sedih dan Saya sering merasa sedih. Responden diminta memilih satu pernyataan dalam setiap kelompok pernyataan.

5. Rational Scale

Skala ini disebut juga skala konsistensi internal. Skala ini berisikan beberapa pernyataan yang direspons Ya-Tidak. Pernyataan-pernyataan ini disekor sesuai dengan kunci jawaban. Setiap pernyataan memiliki kunci jawaban. Kunci jawabannya adalah Ya dan Tidak. Jika kunci jawabannya Ya maka orang yang menjawab Ya diberi skor 1 dan jika menjawab Tidak diberi skor 0. Demikian sebaliknya jika kunci jawabannya Tidak maka skor bagi responden yang menjawab Ya adalah 0 dan yang menjawab Tidak skornya 1. Jadi skor setiap item adalah dikotomi.

Seperti yang dikatakan di atas bahwa skala ini disebut juga dengan konsistensi internal, karena tujuan akhir dari skala ini adalah mencari item-item yang saling memiliki kedekatan satu sama lain. Untuk mencari kedekatan item dengan item yang lain dilakukanlah korelasi antara skor item dan total. Korelasinya adalah point-biserial. Item-item yang memiliki korelasi rendah dan negatif dihapus atau dibuang. Biasanya standar yang dipakai adalah korelasi yang sama atau lebih besar dari 0,2.

Membuat Item

Ada beberapa hal yang harus diperhatikan dalam penyusunan item yaitu:

1. Apakah item yang ditulis bervariasi atau homogen.
2. Berapa rentang tingkat kesulitan yang ingin dicakup?
3. Berapa banyak item awal yang ingin dibuat?
4. Proses kognitif apa yang ingin dicapai?
5. Jenis item seperti apa yang ingin dibuat?

Menguji Item

1. Tingkat kesulitan item

Tingkat kesulitan item hanya untuk item dalam tes kognitif. Rentang indeks tingkat kesulitan item adalah 0,0 sampai 1,0. Semakin kecil indeks semakin sulit item tersebut. Tingkat kesulitan item bisa dikategorikan sebagai berikut:

Kategori	Rentang Indeks
Sangat Sulit	0,0 – 0,20
Sulit	0,21 – 0,40
Sedang	0,41 – 0,60
Mudah	0,61 – 0,80
Sangat Mudah	0,81 – 1,00

2. Reliabilitas item

Reliabilitas item adalah mencari sejauhmana item-item dalam sebuah tes memiliki konsistensi internal yang tinggi. Item-item dalam sebuah tes adalah mengukur perilaku yang sama sehingga item-item tersebut secara logis adalah homogen. Untuk mengetahui homogenitas item adalah dengan mengorelasikan skor item dengan skor total tes. Korelasi yang digunakan adalah korelasi point-biserial untuk item dikotomi dan korelasi product-moment untuk item multikotomi. Item yang dianggap memiliki konsistensi internal dengan tes tersebut adalah yang memiliki korelasi $\geq 0,30$.

3. Validitas item

Kalau reliabilitas item adalah dengan mengorelasikan skor item dengan skor total tes itu sendiri, sedangkan validitas item adalah mengorelasikan skor item tersebut dengan skor kriteria. Skor kriteria ini adalah tes lain yang berkaitan dengan tes yang diuji. Korelasi yang diujikan adalah korelasi point biserial dan korelasi product moment. Misalnya kita sedang menguji skala Need of Achievement. Untuk mengetahui validitas itemnya maka kita korelasikan dengan skor tes Test Anxiety yang memiliki korelasi atau hubungan dengan Need of Achievement. Item yang dianggap memiliki konsistensi internal dengan tes tersebut adalah yang memiliki korelasi $\geq 0,30$.

4. Daya beda item

Daya beda item adalah untuk mengetahui sejauhmana item itu bisa membedakan orang yang memperoleh skor tinggi dan skor rendah pada skor total tes. Secara tidak langsung daya beda item ini bisa dilihat dari korelasi item-total. Semakin indeks daya bedanya tinggi maka semakin tinggi pula korelasi item-totalnya. Rumus daya beda item adalah

$$D = (U-L)/N$$

D : Indeks daya beda

U : Jumlah orang yang menjawab benar pada kelompok atas

L : Jumlah orang yang menjawab benar pada kelompok bawah

N : Jumlah total testee

Rentang indeks daya beda adalah -1,0 sampai 1,0, berikut contoh rentang indeks daya beda.

Item	U	L	D	Komentar
1	49	23	0,26	Item yang bagus dengan tingkat kesulitan tinggi
2	79	19	0,60	Item yang sangat bagus tapi jarang terjadi
3	52	52	0,00	Item yang jelek dan harus direvisi
4	100	0	1,00	Item yang ideal tapi tidak pernah terjadi
5	20	80	-0,60	Item yang sangat buruk dan harus dihapus
6	0	100	-1,00	Item yang secara teoritis yang terburuk

Merevisi Item

Tujuan analisis item adalah mengetahui item-item yang tidak produktif sehingga bisa direvisi. Item-item yang direvisi adalah item-item yang masih ditoleransi kesalahannya. Misalnya item yang korelasi item totalnya 0,1,5 - 0,2. Jadi item yang memiliki korelasi dibawah itu atau negatif adalah item yang harus dibuang tanpa ada pertimbangan untuk direvisi. Setelah item-item direvisi dilakukan sekali uji coba kedua pada sampel lain yang memiliki karakteristik yang sama. Jika perubahan indeks analisis item tidak jauh berbeda dari ujicoba pertama dan kedua maka item-item dianggap memuaskan, tapi jika perubahannya besar maka diperlukan lagi ujicoba yang ketiga atau bahkan yang keempat jika memang dianggap belum memuaskan.

Feedback dari peserta tes

Feedback dari peserta tes merupakan sumber informasi yang penting merevisi sebuah tes. Feedback ini bisa mencakup apa yang dirasakan peserta tes ketika diminta menjawab tes tersebut. Hal-hal ini bisa meliputi:

1. Perilaku penguji
2. Kondisi pengetesan
3. Kejelasan instruksi tes
4. Kenyamanan dalam menggunakan lembar tes
5. Ketersedian waktu yang cukup
6. Tingkat kesulitan tes yang wajar
7. Tingkat penebakan
8. Kecurangan oleh peserta tes