

# PERKULIAHAN 7: EVALUASI PEMBELAJARAN MATEMATIKA

## KUALITAS ALAT EVALUASI

### 1. Pendahuluan

Bertitik tolak dari pendapat Galton, bahwa dalam suatu kelompok individu (siswa) yang tidak dipilih secara khusus memiliki karakteristik tertentu yang frekuensinya berdistribusi normal. Begitu pula kepandaiannya dalam suatu mata pelajaran tertentu. Dengan demikian suatu alat evaluasi yang baik akan mencerminkan kemampuan sebenarnya dari testi yang dievaluasi dan bisa membedakan siswa yang pandai (di atas rata-rata), siswa yang kemampuannya sedang (pada kelompok rata-rata), dan siswa yang kemampuannya kurang (di bawah rata-rata). Sehingga penyebaran skor atau nilai hasil evaluasi tersebut berdistribusi normal.

Untuk mendapatkan hasil evaluasi yang baik tentunya diperlukan alat evaluasi yang kualitasnya baik pula. Alat evaluasi yang baik dapat ditinjau dari hal-hal berikut ini, yaitu: validitas, reliabilitas, daya pembeda, derajat kesukaran, efektifitas option, obyektifitas, dan praktikabilitas.

### 2. Validitas

Suatu alat evaluasi disebut valid (absah atau sah) apabila alat tersebut mampu mengevaluasi apa yang seharusnya dievaluasi. Oleh karena itu keabsahannya tergantung pada sejauh mana ketepatan alat evaluasi itu dalam melaksanakan fungsinya. Dengan demikian suatu alat evaluasi disebut valid jika ia dapat mengevaluasi dengan tepat sesuatu yang dievaluasi itu.

Validitas berdasarkan pelaksanaannya dapat dikelompokkan ke dalam dua jenis, yaitu:

#### 1) Validitas Teoritik

Validitas teoritik atau validitas logik adalah validitas alat evaluasi yang dilakukan berdasarkan pertimbangan (*judgement*) teoritik atau logika yang dilakukan oleh para ahli atau orang yang dianggap ahli. Ada tiga macam validitas yang termasuk ke dalam validitas teoritik ini, yaitu:

a. Validitas Isi

Validitas isi suatu alat evaluasi artinya ketepatan alat tersebut ditinjau dari segi materi yang dievaluasikan, yaitu materi (bahan) yang dipakai sebagai alat evaluasi tersebut yang merupakan sampel representative dari pengetahuan yang harus dikuasai

b. Validitas Muka

Validitas muka suatu alat evaluasi disebut pula validitas bentuk soal (pertanyaan, pernyataan, suruhan) atau validitas tampilan, yaitu keabsahan susunan kalimat atau kata-kata dalam soal sehingga jelas pengertiannya atau tidak menimbulkan tafsiran lain.

c. Validitas Konstruksi Psikologik

Pada umumnya alat evaluasi yang sering menyangkut validitas konstruksi ini berkenaan dengan aspek sikap, kepribadian, motivasi, minat, dan bakat.

2) Validitas Kriterion

Validitas ini diperoleh dengan melalui observasi atau pengalaman yang bersifat empirik, kriterion itu dipergunakan untuk menentukan tinggi-rendahnya koefisien validitas alat evaluasi yang dibuat melalui perhitungan korelasi. Ada dua macam validitas yang termasuk ke dalam validitas kriterion ini, yaitu:

a. Validitas Banding

Validitas banding seringkali disebut validitas bersama. Misalnya alat evaluasi yang diselidiki validitasnya adalah tes matematika buatan guru (kita) dengan menggunakan kriterion nilai rata-rata harian atau nilai tes sumatif yang telah ada, dengan asumsi hasil evaluasi yang digunakan untuk kriterion itu telah mencerminkan kemampuan siswa sebenarnya. Kedua tes tersebut diberikan kepada subyek (siswa) yang sama. Apabila kedua nilai atau skor itu berkorelasi tinggi, maka tes yang kita buat itu memiliki validitas yang tinggi pula.

b. Validitas Ramal

Sebuah alat evaluasi dikatakan memiliki validitas ramal yang baik jika ia mempunyai kemampuan untuk meramalkan hal-hal yang akan terjadi di masa yang akan datang.

Cara menentukan tingkat (indeks) validitas kriterion adalah dengan menghitung koefisien korelasi antara alat evaluasi yang akan diketahui validitasnya

dengan alat ukur lain yang telah dilaksanakan dan diasumsikan telah memiliki validitas yang tinggi (baik). Cara mencari koefisien validitas dapat digunakan 3 macam cara, yaitu: korelasi produk moment memakai simpangan, korelasi produk moment memakai angka kasar, dan korelasi metode rank.

### **3. Reliabilitas**

Reliabilitas suatu alat ukur atau alat evaluasi dimaksudkan sebagai suatu alat yang memberikan hasil yang tetap sama (konsisten, ajeg). Hasil pengukuran itu harus tetap sama (relatif sama) jika pengukurannya diberikan pada subyek yang sama meskipun dilakukan oleh orang yang berbeda, waktu yang berbeda, dan tempat yang berbeda pula. Relatif di sini dimaksudkan tidak tepat sama, tetapi mengalami perubahan yang tak berarti (tidak signifikan) dan bisa diabaikan, Perubahan hasil evaluasi ini disebabkan adanya unsur pengalaman dari peserta tes dan kondisi lainnya.

Untuk soal objektif, menentukan derajat reliabilitas bisa dicari dengan teknik belah dua atau teknik non belah dua. Dengan teknik belah dua ada tiga macam perhitungan, yaitu: Formula Spearman-Brown, Formula Flanagan, dan Formula Rulon. Sedangkan dengan teknik non belah dua ada dua macam perhitungan, yaitu: Kuder Richardson (KR-20 dan KR-21) dan Anava Hoyt. Untuk soal bentuk uraian, menentukan derajat reliabilitas bisa dicari dengan menggunakan rumus Alpha.

### **4. Daya Pembeda**

Pengertian Daya Pembeda (DP) dari sebuah butir soal menyatakan seberapa jauh kemampuan butir soal tersebut mampu membedakan antara testi yang mengetahui jawabannya dengan benar dengan testi yang tidak dapat menjawab soal tersebut (atau testi yang menjawab salah). Dengan perkataan lain daya pembeda sebuah butir soal adalah kemampuan butir soal itu untuk membedakan antara testi (siswa) yang pandai atau berkemampuan tinggi dengan siswa yang bodoh.