

BAB II

ANALISIS KULITAS BUTIR SOAL PADA BANK SOAL FISIKA

A. Konsep Dasar Evaluasi

1. Pengertian Evaluasi

Secara umum, pengertian evaluasi adalah suatu proses untuk menyediakan informasi tentang sejauh mana suatu kegiatan tertentu telah dicapai, bagaimana perbedaan pencapaian itu dengan suatu standar tertentu untuk mengetahui apakah ada selisih di antara keduanya, serta bagaimana manfaat yang telah dikerjakan itu bila dibandingkan dengan harapan-harapan yang ingin diperoleh.

Cross (dalam Sukardi, 2010: 1) mengemukakan bahwa *“evaluation is a process which determines the extent to which objectives have been achieved”* (evaluasi merupakan proses yang menentukan kondisi, dimana suatu tujuan telah dapat dicapai). Menurut Ralph Tyler (dalam Arikunto, 2010: 3) mengatakan bahwa *“Evaluasi merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagian mana tujuan pendidikan sudah tercapai. Jika belum, bagian mana yang belum dan apa sebabnya”*.

Dalam hubungan dengan pengajaran, Grounlund (dalam Purwanto, 2011: 3) mengatakan bahwa *“evaluation a systematic process of determining the extent to which insructional objectives are*

achieved by pupils”(evaluasi adalah suatu proses yang sistematis untuk menentukan sejauh mana tujuan-tujuan pengajaran telah dicapai oleh siswa). Berdasarkan pendapat tersebut, dapat disimpulkan bahwa evaluasi dalam proses belajar mengajar adalah kegiatan sistematis yang dilakukan oleh guru untuk mengetahui apakah tujuan pembelajaran telah dicapai dan sebagai acuan perbaikan kekurangan dalam proses belajar mengajar ke arah yang lebih baik .

Berkaitan dengan evaluasi tidak terlepas dengan penilaian, pengukuran dan tes. Dari keempat istilah ini dalam pembelajaran sering kali dihubungkan dan disamakan. Namun dari keempat istilah tersebut memiliki arti, ruang lingkup maupun fokus yang dinilai berbeda.

Evaluasi lebih luas ruang lingkungnya dari penilaian, sedangkan penilaian lebih berfokus pada aspek tertentu saja yang merupakan bagian dari ruang lingkup tersebut. Jika hal yang ingin dinilai adalah sistem pembelajaran, maka ruang lingkungnya adalah semua komponen pembelajaran, dan istilah yang tepat untuk menilai sistem pembelajaran adalah evaluasi, bukan penilaian. Jika hal yang ingin dinilai satu atau beberapa bagian/komponen pembelajaran misalnya hasil belajar, maka istilah yang tepat adalah penilaian.

Disamping itu, ada juga istilah pengukuran. Untuk dapat menilai dengan tepat, diperlukan adanya suatu alat penilaian berupa pengukuran. Menurut Kerlinger (dalam Purwanto, 2011: 2)

“Pengukuran adalah membandingkan sesuatu yang diukur dengan alat ukurnya dan kemudian menerangkan angka menurut sistem aturan tertentu”. Jadi, Pengukuran merupakan suatu kegiatan untuk membandingkan sesuatu dengan satu ukuran. Kalau evaluasi dan penilaian bersifat kualitatif, maka pengukuran bersifat kuantitatif (skor/angka) yang diperoleh menggunakan alat ukur atau instrumen yang standar. Dalam konteks hasil belajar, alat ukur atau instrumen tersebut dapat berbentuk tes atau no tes.

Untuk mengetahui hasil belajar siswa, diperlukannya suatu kegiatan yang disebut penilaian. Depdikbud tahun 1994 (dalam Arifin, 2009: 4) mengemukakan “Penilaian adalah suatu kegiatan untuk memberikan berbagai informasi secara berkesinambungan dan menyeluruh tentang proses dan hasil yang telah dicapai siswa”. Grondlund (dalam Arifin, 2009: 4) mengartikan “Penilaian adalah suatu proses yang sistematis dari pengumpulan, analisis, dan interpretasi informasi/data untuk menentukan sejauh mana peserta didik telah mencapai tujuan pembelajaran”.

Dapat disimpulkan bahwa, penilaian adalah suatu proses atau kegiatan yang sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil belajar peserta didik dalam rangka membuat keputusan-keputusan berdasarkan kriteria dan pertimbangan tertentu. Keputusan yang dimaksud adalah keputusan tentang peserta

didik, seperti nilai yang akan diberikan atau juga keputusan tentang kenaikan kelas dan kelulusan.

2. Prinsip-Prinsip Evaluasi

Menurut Zulfadrial (2012: 11) adapun prinsip-prinsip penilaian secara umum sebagai berikut:

- a. Valid. Penilaian harus mengukur apa yang seharusnya diukur dengan menggunakan alat yang dapat dipercaya, tepat dan sah.
- b. Mendidik. Penilaian harus memberikan sumbangan positif terhadap pencapaian hasil belajar siswa.
- c. Berorientasi pada kompetensi. Penilaian harus menilai pencapaian kompetensi yang dimaksud dalam kurikulum.
- d. Adil dan Objektif. Penilaian harus adil terhadap semua siswa dan tidak membeda-bedakan latar belakang siswa yang tidak berkaitan dengan pencapaian hasil belajar.
- e. Terbuka. Kriteria penilaian hendaknya terbuka bagi berbagai kalangan sehingga keputusan tentang keberhasilan siswa jelas bagi pihak-pihak berkementingan.
- f. Berkesinambungan. Penilaian dilakukan secara berencana, bertahap, teratur dan terus menerus, dan berkesinambungan untuk memperoleh gambaran tentang perkembangan kemajuan belajar siswa.
- g. Menyeluruh. Penilaian terhadap hasil belajar siswa harus dilaksanakan menyeluruh, utuh dan tuntas yang mencakup aspek

kognitif, psikomotorik dan afektif serta berdasarkan pada berbagai teknik dan prosedur penilaian dengan berbagai bukti hasil belajar siswa.

- h. Bermakna. Penilaian hendaknya mudah dipahami dan bisa ditinjau lanjuti oleh pihak-pihak yang berkepentingan.

3. Tujuan Evaluasi

Secara umum, evaluasi bertujuan untuk mengukur kemampuan siswa setelah mengikuti proses belajar mengajar. Sukardi (2010: 9) mengemukakan bahwa tujuan evaluasi adalah sebagai berikut:

- a. Menilai ketercapaian (*attainment*) tujuan.
- b. Mengukur macam-macam aspek belajar yang bervariasi.
- c. Sebagai saran (*means*) untuk mengetahui apa yang siswa telah ketahui.
- d. Memotivasi belajar siswa.
- e. Menyediakan informasi untuk tujuan bimbingan dan konseling. Menjadikan hasil evaluasi sebagai dasar perubahan kurikulum.

Dapat disimpulkan bahwa tujuan evaluasi adalah untuk mengukur penyapaian tujuan, mengetahui kemampuan siswa dalam hal-hal tertentu, menentukan layak tidaknya seorang siswa dinyatakan naik kelas atau lulus dan untuk memberikan umpan balik bagi guru mengenai kegiatan belajar mengajar yang dilakukan.

B. Tes Sumatif

Sudaryono (2012: 40) mengungkapkan bahwa dilihat dari segi aspek hasil belajar yang dievaluasi, maka kita akan melihat adanya evaluasi yang berhubungan dengan hasil belajar kognitif, afektif dan psikomotorik. Selanjutnya Zulfadrial (2012: 33) juga menjelaskan bahwa evaluasi hasil belajar disekolah meliputi tiga domain yaitu domain *cognitive*, *affective* dan *psychomotor*.

Domain *cognitive* adalah kemampuan berfikir, domain *affective* adalah kemampuan dalam sikap dan domain *psychomotor* adalah kemampuan motorik. Dengan ranah tujuan pembelajaran sebagaimana di kemukakan oleh Bloom adalah sebagai berikut *cognitive domain* meliputi: (1) pengetahuan, (2) pemahaman, (3) penerapan, (4) analisis, (5) sintesis dan (6) evaluasi. *Affective domain* meliputi: (1) kemampuan menerima/penerima, (2) kemampuan menanggapi/pemberian respon, (3) berkeyakinan/penghargaan, (4) penerapan karya/pengorganisasian, (5) ketekunan/ketelitian/karakterisasi. *Psikomotorik Domain* meliputi: (1) Persepsi, (2) kesiapan, (3) respon terbimbing, (4) mekanisme, (5) respon yang kompleks, (6) adaptasi dan (7) orijinasi.

Penilaian aspek kognitif dalam bentuk tes yaitu tes lisan dan tes tertulis atau tes hasil belajar. Penilaian efektif penilaian dalam bentuk tes sikap, minat, motivasi, nilai dan moral. Penilaian psikomotorik penilaian dalam bentuk unjuk kerja atau perbuatan. Salah satu penilaian dalam hasil

belajar yaitu penilaian sumatif. Kata sumatif berasal dari bahasa Inggris yaitu “sum” yang artinya jumlah atau total. Tes sumatif yang dimaksud sebagai tes yang digunakan untuk mengetahui penguasaan siswa atas semua jumlah materi yang disampaikan dalam satuan waktu tertentu seperti catur wulan atau semester (Purwanto, 2010: 68). Setelah semua materi selesai disampaikan, maka evaluasi dilakukan atas perubahan perilaku yang terbentuk pada siswa setelah memperoleh semua materi pelajaran. Evaluasi dilakukan berdasarkan hasil pengukuran tes sumatif. Dalam praktik pengajaran tes sumatif dikenal sebagai ujian akhir semester atau catur wulan tergantung satuan waktu yang digunakan untuk menyelesaikan materi.

C. Analisis Kualitas Butir Soal

Menurut Kamus Besar Bahasa Indonesia (2016), analisis adalah penguraian suatu pokok atas berbagai bagiannya dan penelaahan bagian itu sendiri serta hubungan antarbagian untuk memperoleh pengertian yang tepat dan pemahaman arti keseluruhan. Kualitas adalah tingkat baik buruknya sesuatu. Jadi yang dimaksud dengan analisis kualitas butir soal merupakan penelaahan yang dilakukan untuk mengetahui baik buruknya butir soal.

Menurut Arikunto (2010: 205) analisis butir soal adalah suatu prosedur yang sistematis yang akan memberikan informasi sangat khusus terhadap butir tes yang di susun. Arifin (2009: 246) menyatakan, “Analisis

kualitas tes merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu tes, baik tes secara keseluruhan maupun butir soal yang menjadi bagian dari tes tersebut". Menurut Daryanto (2007: 179) mengemukakan analisis soal bertujuan untuk mengadakan identifikasi soal-soal yang baik, kurang baik dan soal yang jelek. Dengan analisis soal dapat diperoleh informasi tentang kejelekan sebuah soal dan petunjuk untuk mengadakan perbaikan.

Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan guru untuk meningkatkan mutu soal yang telah ditulis. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban siswa untuk membuat keputusan tentang setiap penilaian. Soal yang bermutu adalah soal yang dapat memberikan informasi setepat-tepanya sesuai dengan tujuannya diantaranya dapat menentukan peserta didik mana yang sudah atau belum menguasai materi yang diajarkan guru.

Menurut Djauarsih (2010: 1), ada dua teknik yang dapat digunakan dalam penelaah butir soal, yaitu penelaah soal secara kualitatif dan secara kuantitatif. Adapun teknik terbaik adalah dengan mengabungkan keduanya.

1. Analisis Kualitas Butir Soal Secara Kualitatif

Pada prinsipnya analisis butir soal secara kualitatif dilaksanakan berdasarkan kaidah penulisan soal (tes tertulis, perbuatan, dan sikap). Penelaahan ini biasanya dilakukan sebelum soal

digunakan atau diujikan. Sudaryono (2012: 138) mengungkapkan bahwa analisis kualitatif terhadap suatu soal berdasarkan kriteria yang telah ditentukan, dalam hal ini kriteria yang dimaksud adalah kriteria materi, konstruksi dan bahasa.

a) Aspek Isi

Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu sejajar dengan materi atau isi pelajaran yang diberikan (Arikunto, 2008: 89). Sedangkan Validitas isi menurut Sudaryono (2012: 140) mengungkapkan bahwa validitas isi dari suatu tes hasil belajar adalah validitas yang diperoleh setelah dilakukan penganalisisan, penelusuran atau pengujian terhadap isi yang dikandung dalam tes hasil belajar tersebut.

Validitas isi adalah validitas yang dilihat dari segi isi tes itu sendiri sebagai alat pengukuran hasil belajar yaitu: sejauh mana tes hasil belajar sebagai alat pengukuran hasil belajar peserta didik, isinya telah dapat mewakili secara representatif terhadap keseluruhan materi atau bahan pelajaran yang seharusnya ditekankan. Instrumen dapat dikatakan memenuhi validitas isi apabila materi yang diukur tersebut sesuai dengan materi yang tertuang dalam kurikulum.

b) Aspek konstruksi

Menurut Arikunto (2008: 90) sebuah tes dikatakan memiliki validitas konstruk jika butir soal yang membangun tes tersebut mengukur setiap aspek berfikir seperti yang dirumuskan dalam indikator. Tes hasil belajar baru dapat dikatakan memiliki validitas susunan apabila butir-butir soal yang membangun tes tersebut benar-benar dapat dengan secara tepat mengukur aspek-aspek berfikir (seperti: aspek kognitif, aspek efektif, aspek psikomotorik dan sebagainya) sebagaimana telah ditentukan dalam tujuan instruksional khusus.

c) Aspek bahasa

Menurut Sapranata (dalam Rahmadhani, 2014: 424), analisis bahasa yang dimaksudkan adalah penelaah soal yang berkaitan dengan penggunaan Bahasa Indonesia yang baik dan benar menurut EYD.

Menurut Guion dalam Sudaryono (2012: 140), validitas isi sangat tergantung kepada dua hal yaitu tes itu sendiri dan proses yang mempengaruhi dalam merespon tes. Djanuarsih mengemukakan bahwa (2012: 5) ada beberapa teknik yang dapat digunakan untuk menganalisis butir soal secara kualitatif, diantaranya adalah teknik moderator dan teknik panel.

Teknik moderator merupakan teknik berdiskusi yang didalamnya terdapat satu orang sebagai penengah. Berdasarkan teknik ini, setiap butir soal didiskusikan secara bersama-sama

dengan beberapa ahli seperti guru yang mengajarkan materi, ahli materi, penyusun/pengembang kurikulum, ahli penilaian, ahli bahasa berlatar belakang psikologi.

Teknik panel merupakan suatu teknik menelaah butir soal yang setiap butir soalnya ditelaah berdasarkan kaidah penulisan butir soalnya ditelaah berdasarkan kaidah penulisan butir soal, yaitu ditelaah dari segi materi, konstruksi, bahasa/budaya, kebenaran kunci jawaban/pedoman penskorannya yang dilakukan oleh beberapa penelaah. Caranya adalah beberapa penelaah diberikan: butir-butir soal yang akan ditelaah, format penelaahan dan pedoman penilaian/penelaahannya. Pada tahap awal para penelaah diberikan pengarahan, kemudian tahap berikutnya para penelaah bekerja sendiri-sendiri di tempat yang tidak sama. Dalam menganalisis butir soal secara kualitatif, penggunaan format penelaahan soal akan sangat membantu dan mempermudah prosedur pelaksanaannya.

2. Analisis Kualitas Butir Soal Secara Kuantitatif

Penelaah soal secara kuantitatif maksudnya adalah penelaahan butir soal didasarkan pada data empirik dari butir soal yang bersangkutan. Data empirik ini diperoleh dari soal yang telah diujikan. Menurut Daryanto (2007: 179) mengemukakan bahwa kapan soal itu dikatakan baik, ketika mencakup taraf kesukaran dengan soal yang

tidak terlalu mudah atau tidak terlalu sukar. Daya pembeda adalah kemampuan soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah), dan pola jawaban yang baik adalah yang dapat berfungsi mengecoh siswa untuk menjawab soal dengan benar.

Menurut Nurkencana dan Sunartana (1986: 127) baik buruknya suatu evaluasi dapat ditinjau dari beberapa segi, yaitu validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Menurut Pratiwi (2013) Alat evaluasi yang berkualitas itu harus memiliki lima karakteristik yaitu daya beda, tingkat kesukaran, efektivitas pengecoh, validitas dan reliabilitas. Analisis kualitas butir soal pada penelitian ini meliputi:

1. Taraf Kesukaran

Tingkat kesukaran adalah pengukuran seberapa besar derajat kesukaran suatu soal. Jika soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya (Daryanto, 2007: 179).

Tingkat kesukaran merupakan salah satu ciri tes yang perlu diperhatikan, karena tingkat kesukaran menunjukkan seberapa

sukar atau mudahnya butir-butir tes secara keseluruhan yang telah diselenggarakan. Perhitungan tingkat kesukaran soal adalah pengukuran sebesar derajat kesukran suatu soal. jika soal memiliki tingkat kesukaran seimbang (proporsional), maka dapat dikatakan bahwa soal tersebut baik.

Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran (*difficulty Index*) . besarnya Indeks kesukran antara 0,00 sampai dengan 1,0. Indeks kesukaran ini menunjukkan taraf kesukaran soal . soal dengan indeks kesukaran 0,0 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,0 menunjukkan bahwa soalnya terlalu mudah. Di dalam Istilah evaluasi, indeks kesukaran ini diberikan simbol P (p besar), singkatan dari kata “proporsi”. Dengan demikian maka soal dengan $P = 0,70$ lebih mudah jika dibandingkan dengan $P = 0,20$. Sebaliknya soal dengan $P = 0,30$ lebih sukar dari pada soal dengan $P = 0,80$.

Rumus mencari P adalah

$$P = \frac{B}{JS} \quad \dots(2.1)$$

Di mana:

P = indeks kesukaran

B = banyaknya siswa yang menjawab soal itu dengan betul

JS = jumlah seluruh siswa peserta tes (Arikunto,2005:208).

Menurut Purwanto (2009: 99) menggunakan rumus

sebagai berikut dengan kriteria tingkat kesukaran butir soal

sebagai berikut:

Tabel 2.1
Kriteria Tingkat Kesukaran Butir Soal

Rentang TK	Kategori
0,00 – 0,32	Sukar
0,33 – 0,66	Sedang
0,67 – 1,00	Mudah

2. Daya Pembeda

Daryanto (2007: 183) menyatakan, “Daya pembeda soal adalah kemampuan sesuatu soal untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang bodoh (berkemampuan rendah). Semakin tinggi koefisien daya pembeda suatu butir soal, semakin mampu butir soal tersebut membedakan antara peserta didik yang menguasai kompetensi dengan peserta didik yang kurang menguasai kompetensi peserta didik.

Dengan melakukan analisis tes, dapat membantu dalam mengidentifikasi butir-butir soal yang kurang baik, sehingga memperoleh informasi yang dapat digunakan untuk menyempurnakan soal soal untuk kepentingan selanjutnya. Manfaat terbesar dari analisis tes ini adalah guru diharapkan semakin memahami bagaimana wujud tes yang baik dan bagaimana butir soal yang baik, sehingga pada akhirnya guru semakin terampil menyusun soal tes dengan baik dan berkualitas.

Salah satu ciri butir soal yang baik adalah yang mampu membedakan antara kelompok atas (yang mampu) dan kelompok bawah (kurang mampu), karena itu butir tes harus diketahui daya

bedanya. Siswa yang berkemampuan tinggi adalah siswa yang mempunyai rata-rata skor paling baik. siswa yang termasuk kelompok rendah adalah siswa yang mempunyai rata-rata skor yang rendah.

Bagi suatu soal yang dapat dijawab benar oleh siswa pandai maupun siswa kurang pandai, maka soal itu tidak baik karena tidak mempunyai daya pembeda. Demikian pula jika semua siswa baik pandai maupun kurang pandai tidak dapat menjawab dengan benar. Soal itu tidak baik juga karena tidak mempunyai daya pembeda.

Soal yang baik adalah soal yang dapat dijawab benar oleh siswa-siswa yang pandai saja. Angka yang menunjukkan besarnya daya pembeda disebut indeks deskriminasi, disingkat D (d besar). Seperti halnya indeks kesukaran, indeks deskriminasi (daya pembeda) ini berkisar antara 0,00 sampai 1,00. Menurut Arikunto (2007: 190) rumus untuk menentukan indeks deskriminasi adalah:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B \quad \dots(2.2)$$

Keterangan :

J = jumlah peserta tes

J_A = banyaknya peserta kelompok atas

J_B = banyaknya peserta kelompok bawah

$B_A = \frac{B_A}{J_A}$ = banyaknya peserta kelompok atas yang menjawab soal itu dengan benar

$P_A = \frac{B_B}{J_B}$ = banyaknya peserta kelompok bawah yang menjawab soal itu dengan benar.

Dengan klasifikasi daya pembeda menurut Arikunto (2010: 218) sebagai berikut:

Tabel 2.2
Klasifikasi Daya Pembeda

Interval	Kriteria
0,00 – 0,19	Jelek
0,20 – 0,39	Cukup
0,40 – 0,69	Baik
0,70 – 1,00	baik sekali
Negatif	Sangat jelek

Sebuah butir *THB* yang baik adalah butir soal yang mempunyai *DB* positif dan Signifikan. Purwanto (2009: 105) mengemukakan bahwa :

“*DB* akan positif apabila jumlah siswa kelompok atas yang dapat menjawab dengan benar lebih banyak daripada jumlah siswa kelompok bawah. *DB* yang signifikan dimaksudkan sebagai mempunyai indeks minimal +0,30 yang artinya pada butir yang baik jumlah siswa kelompok atas yang dapat menjawab benar minimal 30% lebih banyak daripada jumlah siswa kelompok bawah yang dapat menjawab benar”.

3. Efektivitas Pengecoh

Pengecoh juga dikenal dengan istilah penyesat atau penggoda adalah pilihan jawaban yang merupakan jawaban dari soal. analisis butir juga dilakukan dengan memperhatikan pengecoh. Pengecoh bukan hanya sekedar pelengkap pilihan. Pengecoh diadakan untuk menyesatkan siswa agar tidak memilih kunci jawaban.

Menurut Daryanto (2007: 193) pengecoh dapat dikatakan berfungsi baik jika paling sedikit dipilih oleh 5% pengikut tes. Menurut Purwanto (2009: 108) pengecoh yang sama sekali tidak dipilih tidak dapat melakukan fungsinya sebagai pengecoh karena terlalu menyolok dan dimengerti oleh semua siswa sebagai pengecoh soal. Pengecoh yang berdasarkan hasil uji coba tidak efektif direkomendasikan untuk diganti dengan pengecoh yang lebih menarik.

4. Validitas

Validitas dapat berkenaan dengan ketepatan alat penilaian terhadap konsep yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai. Scarvia B. Anderson dkk (dalam Arikunto, 2005: 64) mengemukakan “*A test is valid if it measures what it purpose to measure*” (sebuah tes dikatakan valid apabila tes tersebut mengukur apa yang hendak di ukur. Dalam bahasa indonesia “valid” disebut dengan istilah sah. Arikunto (2005: 67) menjelaskan ada empat bentuk validitas yaitu: validitas isi, validitas konstruk, validitas ada sekarang dan validitas prediksi.

a. Validitas Isi (*Content Validity*)

Validitas isi adalah suatu validitas yang menunjukkan sampai dimana isi suatu tes atau alat pengukur mencerminkan hal-hal yang mau di ukur atau yang diteskan. Khusus tes hasil

belajar yang telah direncanakan dengan baik antara lain tahap merumuskan tujuan instruksional dan merincikan bahan pelajaran, maka kedua tahap tersebut mempunyai kaitan dengan validitas isi suatu tes hasil belajar.

Alat tes yang dianggap layak dan dapat dipertanggungjawabkan validitas isinya apabila dalam penyusunannya berdasarkan tabel kisi-kisi pembuatan soal. validitas isi hendaknya merujuk pada kesesuaian antara butir-butir soal dengan kompetensi dasar dan standar kompeteninya.

Secara sederhana dapat dikatakan bahwa tes yang disusun tidak boleh keluar dari standar kompetensi mata pelajaran yang ada didalam kurikulum. Disini sangatlah jelas betapa pentingnya kedudukan suatu perencanaan oleh seorang guru, sehingga mampu menunjukkan mutu taraf validitas isi yang dibuatnya.

b. Validitas Konstruk Atau Konsep (*Concept Or Construct Validity*)

Yang dimaksud validitas konstruk adalah suatu validitas yang menunjukkan sampai dimana isi suatu tes atau alat pengukuran sesuai dengan konsep yang seharusnya menjadi isi tes atau alat pengukuran tersebut atau konstruksi teoritis yang mendasari disusunnya tes atau alat pengukur tersebut.

Apabila isi item-item yang merupakan suatu kesatuan suatu tes benar-benar sesuai dengan suatu konsep atau konstruksi yang seharusnya menjadi isinya yaitu mengukur setiap aspek berfikir (ingatan, pemahaman dan aplikasi) seperti yang disebutkan pada indikator dalam kisi-kisi, maka diaktakan tes tersebut memiliki validitas konsep yang tinggi.

c. Validitas kriteria (*Criterion-Related Validity*)

Yang dimaksud validitas kriteria adalah suatu validitas yang memperhatikan hubungan yang ada antara tes atau alat pengukur dengan pengukur lain yang berfungsi sebagai kriteria tau pembanding. suatu kriteria yang baik harus memenuhi syarat-syarat seperti relevan, reliabel bebas dari kesalahan pengukuran dan mudah diperoleh yang kemudian hasil pegukuran dari suatu tes yang akan diperiksa teraf validitasnya diperbandingkan dengan suatu kriteria.

Pengertian umum untuk validitas butir soal adalah sebuah butir soal dikatakan valid apabila mempunyai dukungan besar terhadap skor total. Skor pada butir soal menyebabkan total menjadi tinggi atau rendah. Dengan kata lain, bahwa sebuah butir soal memiliki validitas yang tinggi jika skor pada butir soal mempunyai kesejajaran dengan skor total.

5. Reliabilitas

Syarat lain yang juga penting bagi suatu instrumen evaluasi adalah terpenuhinya syarat kedua selain validitas yaitu reliabilitas. Reliabilitas merupakan ketetapan atau keajegan alat tersebut dalam menilai apa yang dinilai. Artinya, kapan pun alat penilaian tersebut digunakan akan memberikan hasil yang relatif sama. Semakin reliabel suatu tes, semakin yakin kita dapat menyatakan bahwa dalam hasil suatu tes mempunyai hasil yang sama dan bisa dipakai di suatu tempat sekolah, ketika dilakukan tes kembali (Sukardi, 2010: 43).

Menurut Arikunto (2005: 90), ada 3 cara yang dapat digunakan untuk mencari taraf reliabilitas suatu tes, yaitu:

a) Metode bentuk paralel (*equivalent*)

Tes paralel atau tes *equivalent* adalah dua buah tes yang mempunyai kesamaan tujuan, tingkat kesukaran, dan susunan, tetapi butir-butir soalnya berbeda. Dua buah tes, diteskan kepada kelompok siswa yang sama, kemudian hasilnya dikorelasikan. Koefisien korelasi dari kedua hasil tes inilah yang menunjukkan koefisien reliabilitas tes. Jika koefisiennya tinggi maka tes tersebut sudah reliabel dan dapat digunakan sebagai alat pengetes yang terandal.

b) Metode tes ulang

Dalam menggunakan cara ini, pengetes hanya memiliki satu seri tes tetapi dicobakan dua kali. Kemudian hasil dari kedua kali tes tersebut dihitung korelasinya.

c) Metode belah dua

Dalam teknik ini, pengetes hanya menggunakan sebuah tes dan dicobakan satu kali. Tes yang diberikan dibelah menjadi dua bagian diberikan skor secara terpisah. Ada dua prosedur yang digunakan untuk membelah dua sesuatu tes, yaitu: prosedur ganjil-genap dan prosedur secara random.

Menurut Sudijono (2007: 213) pada penentuan reliabilitas tes hasil belajar bentuk objektif dapat digunakan dengan menggunakan tiga macam pendekatan, yaitu: (1) pendekatan Single Test-Single Trial (*Singel Test-Singel Trial Method*), dilakukan dengan jalan melakukan pengukuran terhadap satu kelompok subjek, dimana pengukuran itu dilakukan dengan hanya menggunakan satu jenis alat pengukuran dan pelaksanaan pengukuran hanya dilakukan sebanyak satu kali saja. (2) pendekatan tes-retest (*Singel Test-Double Trial Method*), penentuan reliabilitas dilakukan dengan menggunakan teknik ulangan, dimana siswa hanya menggunakan satu seri tes, tapi percobaannya dilakukan sebanyak dua kali. (3) pendekatan

Alternate Form (*Double Test-double Trial Method*), yaitu skor-skor yang diperoleh dari kedua seri tes dicari korelasinya. Apabila terdapat korelasi positif yang signifikan, maka dapat dikatakan bahwa tes hasil belajar tersebut dapat dikatakan reliabel.

Sudijono (2007: 213) mengemukakan dengan menggunakan pendekatan singel test-singel trial, maka tinggi rendahnya reliabilitas tes hasil belajar bentuk objektif dapat diketahui dengan melihat besar kecilnya koefisien reliabilitas tes. Adapun untuk menghitung reliabilitas dapat digunakan lima jenis formula, yaitu: (1) formula Spearman-Brown, (2) formula Flanagan, (3) formula Rulon, (4) formula Kuder-Richardson dan (5) Formula C. Hyot.

Formula Spearman-Brown, formula Flanagan dan formula Rulon penentuan reliabilitas tes objektif dilakukan dengan membelah dua tes, sehingga sering diaktakan bahwa ketiga formula itu menggunakan teknik belah dua (*split-half technique*). Penentuan reliabilitas tes banyak ditemukan menggunakan formula Kuder-Richardson dilakukan dengan jalan penganalisisan secara langsung terhadap skor-skor item tes hasil belajar yang bersangkutan, karena apabila dilakukan pembelahan tes menjadi dua belahan maka bisa terjadi koefisien diperoleh berbeda-beda besarnya.

Formula yang dilakukan oleh Kuder dan Richardson ada dua buah, yang masing-masing diberi kode *KR-20* dan *KR-21*, menurut Sukadi (2010: 49) *KR-20* digunakan apabila item tes menggunakan dua pilihan jawaban saja misalnya betul dan salah (B-S). Sedangkan *KR-21* digunakan untuk item tes yang sistematisanya menggunakan pilihan ganda misalnya pilihan ganda empat jawaban, tiga jawaban dan sebagainya. Berikut rumus *KR-20* dan *KR-21*:

a. Rumus *KR-20*:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S_t^2 - \sum p_i q_i}{S_t^2} \right) \quad \dots(2.3)$$

di mana :

- r_{11} = koefisien reliabilitas tes
- n = banyaknya butir item
- S_t^2 = varian total
- p_i = proporsi teste yang menjawab dengan betul butir item yang bersangkutan
- q_i = proporsie testee yang jawabannya salah
($q_i = 1 - p_i$)

b. Rumus *KR-21*:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(1 - \frac{M_t (n - M_t)}{(n)(S_t^2)} \right) \quad \dots(2.4)$$

di mana:

- r_{11} = koefisien reliabilitas tes
- n = banyaknya butir item
- S_t^2 = varian total
- M_t = mean total (rata-rata hitung dari skor total)

Dengan interpretasi nilai r_{11} mengacu pada pendapat Guilford (dalam Jihad dan Haris, 2008: 181) adalah sebagai berikut:

Tabel 2.3
Interpretasi Nilai Reliabilitas

Interval	Kriteria
$\leq 0,20$	reliabilitas sangat rendah
0,20 – 0,40	reliabilitas renda
0,40 – 0,70	reliabilitas sedang
0,70 – 0,90	reliabilitas tinggi
0,90 – 1,00	reliabilitas sangat tinggi

