

BAB II

LANDASAN TEORI

A. Kajian Pustaka

Pada hakikatnya urgensi kajian penelitian adalah sebagai bahan auto kritik terhadap penelitian yang ada, mengenai kelebihan maupun kekurangannya, sekaligus sebagai bahan perbandingan terhadap kajian yang terdahulu. Dan untuk menghindari terjadinya pengulangan hasil temuan yang membahas permasalahan yang sama dan hampir sama dari seseorang, baik dalam bentuk skripsi, buku dan dalam bentuk tulisan lainnya maka penulis akan memaparkan bentuk tulisan yang sudah ada.

Dalam penelitian ini, peneliti mendapati karya ilmiah yang berupa penelitian tentang analisis soal di sekolah yang peneliti anggap mempunyai relevansi dengan penelitian yang peneliti lakukan. Ada beberapa bentuk tulisan penelitian yang akan penulis paparkan.

1. Skripsi yang disusun oleh Andi Ashari (1706500926)), Mahasiswa Fakultas Keguruan dan Ilmu Pendidikan Universitas Pancasakti Tegal dengan judul “Uji Validitas Konstruk Soal Ulangan Akhir Semester I Mata Pelajaran Matematika Kelas VII SMP Negeri Se-Kecamatan Balapulang Tahun Pelajaran 2009/2010”. Tujuan penelitian yang dilakukan oleh Andi Ashari adalah untuk mengetahui apakah soal Ulangan Akhir Semester (UAS) I mata pelajaran matematika kelas VII SMP Negeri se-kecamatan Balapulang tahun pelajaran 2009/2010 memenuhi validitas isi dan validitas konstruk. Simpulan dari penelitian ini adalah: Soal Ulangan Akhir Semester (UAS) I mata pelajaran matematika kelas VII SMP Negeri se-kecamatan Balapulang tahun pelajaran 2009/2010 memenuhi validitas isi dan validitas konstruk.

Judul dari skripsi peneliti adalah Analisis Soal Akhir Semester I Kelas X SMA Negeri Banyumas Mata Pelajaran Matematika Tahun Pelajaran 2011/2012. Antara judul skripsi ini dengan judul skripsi

peneliti ada persamaannya, yaitu sama-sama membahas tentang soal ulangan akhir semester I.

Perbedaan skripsi ini dengan skripsi peneliti yaitu terletak pada soal, kelas, sekolah, tahun pelajaran dan tujuan penelitian yang diambil. Dalam skripsi ini tujuannya hanya untuk mengetahui validitas isi dan validitas konstruk saja, sedangkan tujuan skripsi peneliti adalah untuk mengetahui validitas isi, validitas konstruksi, reliabilitas, tingkat kesukaran dan daya pembeda.

2. Skripsi yang disusun oleh Dini Kurniawati (209821420929), Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Malang dengan judul "Validasi Soal Ulangan Akhir Semester (UAS) Mata Pelajaran Geografi Semester 2 Kelas X di SMA Negeri Kepanjen Kabupaten Malang Tahun Ajaran 2007/2008". Penelitian ini bertujuan untuk mengetahui validitas soal termasuk di dalamnya validitas kurikuler dan validitas item. Tujuan selanjutnya adalah untuk mengetahui tingkat kesukaran (TK), daya beda (DB), reliabilitas soal, dan untuk mengetahui kekuatan masing-masing opsi dalam soal (analisis opsi) khusus untuk mata pelajaran geografi. Hasil penelitian menunjukkan soal UAS geografi ini memiliki validitas kurikuler yang baik dan validitas item yang jelek. Proporsi distribusi tingkat kesukaran soal geografi adalah jelek sedangkan daya beda soal adalah cukup baik. Soal UAS Geografi ini memiliki hasil reliabilitas yang tinggi yaitu dengan hasil perhitungan 0,7423 baik untuk taraf kepercayaan 0,05 maupun 0,01. Pada analisis opsi diperoleh hasil yaitu sebagian besar soal memiliki distraktor yang jelek, artinya dari 4 distraktor yang disiapkan ada rata-rata 3 distraktor yang tidak berfungsi sama sekali dalam mengecoh peserta didik.

Terdapat persamaan antara judul skripsi ini dengan judul skripsi peneliti, yaitu sama-sama membahas soal ulangan akhir semester kelas X. Judul skripsi ini menggunakan kata validasi, sedangkan judul skripsi peneliti menggunakan kata analisis. Perbedaannya terletak pada soal, semester, mata pelajaran, sekolah dan tahun pelajaran yang diambil.

Untuk tujuan penelitian hampir sama hanya saja tujuan dalam skripsi ini terdapat satu tujuan lagi yaitu untuk mengetahui analisi opsi.

B. Kerangka Teoritik

1. Ulangan Akhir Semester

Dalam Peraturan Menteri Pendidikan Nasional Nomor 20 tahun 2007 tentang standar penilaian pendidikan disebutkan bahwa ulangan adalah proses yang dilakukan untuk mengukur pencapaian kompetensi peserta didik secara berkelanjutan dalam proses pembelajaran, untuk memantau kemajuan, melakukan perbaikan pembelajaran, dan menentukan keberhasilan belajar peserta didik. Sedangkan ulangan akhir semester adalah kegiatan yang dilakukan oleh pendidik untuk mengukur pencapaian kompetensi peserta didik di akhir semester. Mengukur di sini berarti menilai, yang dilakukan dengan jalan menguji pencapaian kompetensi peserta didik dalam bentuk tes hasil belajar. Ulangan umum semester pertama soalnya diambil dari materi semester pertama.¹ Cakupan ulangan meliputi seluruh indikator yang merepresentasikan semua kompetensi dasar pada semester tersebut. Ulangan akhir semester ini bertujuan untuk mengukur kompetensi lulusan pada mata pelajaran tertentu. Dalam penelitian ini adalah mata pelajaran matematika.

2. Analisis Tes Hasil Belajar

Tes hasil belajar merupakan tes penguasaan, karena tes ini mengukur penguasaan peserta didik terhadap materi yang diajarkan oleh guru atau dipelajari oleh peserta didik.² Tes diujikan setelah peserta didik memperoleh materi yang sebelumnya telah diajarkan oleh guru. Tes hasil belajar dilakukan untuk mengukur hasil belajar yakni sejauh mana perubahan perilaku yang diinginkan dalam tujuan pembelajaran telah dapat dicapai oleh peserta didik.

¹ E. Mulyasa, *Kurikulum Tingkat Satuan Pendidikan Sebuah Panduan Praktis* (Bandung: PT Remaja Rosdakarya, 2007), hlm. 259.

² Purwanto, *Evaluasi Hasil Belajar* (Yogyakarta: Pustaka Pelajar, 2011), hlm. 66.

Sebuah tes hasil belajar dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, diantaranya adalah:³

- a. Validitas
- b. Reliabilitas
- c. Objektivitas
- d. Praktikabilitas
- e. Ekonomis

Dalam penelitian ini akan dibahas tentang analisis kualitas tes hasil belajar yaitu validitas dan reliabilitas dan analisis butir soal tes hasil belajar yaitu tingkat kesukaran dan daya pembeda.

a. Analisis Kualitas Tes Hasil Belajar

1) Validitas

Validitas sering diartikan kesahihan.⁴ Validitas adalah kualitas yang menunjukkan hubungan antara suatu pengukuran (diagnosis) dengan arti atau tujuan kriteria belajar atau tingkah laku.⁵ Suatu alat ukur disebut memiliki validitas bilamana alat ukur tersebut isinya layak mengukur objek yang seharusnya diukur dan sesuai dengan kriteria tertentu. Artinya adanya kesesuaian antara alat ukur dengan fungsi pengukuran dan sasaran pengukuran.

Validitas suatu instrumen evaluasi, tidak lain adalah derajat yang menunjukkan di mana suatu tes mengukur apa yang hendak

³ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2010), hlm. 57.

⁴ M. Chabib Thoha, *Teknik Evaluasi Pendidikan* (Jakarta: PT RajaGrafindo Persada, 2001), hlm. 109.

⁵ M. Ngalim Purwanto, *Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran* (Bandung: PT Remaja Rosdakarya, 2002), hlm.137.

diukur.⁶ Validitas suatu instrumen evaluasi mempunyai beberapa makna penting diantaranya seperti berikut.⁷

- a) Validitas berhubungan dengan ketepatan interpretasi hasil tes atau instrumen evaluasi untuk grup individual dan bukan instrumen itu sendiri.
- b) Validitas diartikan sebagai derajat yang menunjukkan kategori yang bisa mencakup kategori rendah, menengah dan tinggi.
- c) Prinsip suatu tes valid, tidak universal. Validitas suatu tes yang perlu diperhatikan oleh para peneliti adalah bahwa ia hanya valid untuk suatu tujuan tertentu saja. Tes valid untuk bidang studi matematika belum tentu valid untuk bidang yang lain misalnya bidang mekanika teknik.

Validitas memiliki beberapa karakteristik, antara lain:⁸

- a) Menunjuk kepada hasil dari penggunaan instrumen tersebut bukan pada instrumennya.
- b) Menunjukkan suatu derajat atau tingkatan, validitasnya tinggi, sedang atau rendah, bukan valid atau tidak valid.
- c) Tidak berlaku umum. Suatu tes matematika menunjukkan validitas tinggi untuk mengukur keterampilan menghitung, tetapi hanya sedang dalam mengukur kemampuan berpikir matematis, bahkan rendah dalam memprediksi keberhasilan dalam matematika untuk masa yang akan datang.

Ada dua unsur penting dalam validitas. Pertama, validitas menunjukkan suatu derajat, ada yang sempurna, ada yang sedang dan

⁶ Singarimbun dan Sofian Effendi, *Metode Penelitian Survai* (Jakarta: LP3ES, 2011), hlm. 122.

⁷ Sukardi, *Evaluasi Pendidikan Prinsip dan Operasionalnya* (Jakarta: PT Bumi Aksara, 2009), hlm. 31.

⁸ Nana Syaodih Sukmadinata, *Metode Penelitian Pendidikan* (Bandung: PT Remaja Rosdakarya, 2010), hlm. 228-229.

ada yang rendah. Kedua, validitas selalu dihubungkan dengan suatu putusan atau tujuan yang spesifik. Sebagaimana pendapat R.L. Thorndike dan H.P. Hagen bahwa “*validity is always in relation to a specific decision or use*”. Sementara itu, Gronlund mengemukakan ada tiga faktor yang mempengaruhi validitas hasil tes, yaitu:⁹

- a) Faktor instrumen evaluasi
- b) Faktor administrasi evaluasi dan penskoran
- c) Faktor dari jawaban peserta didik

Banyak faktor yang dapat mempengaruhi hasil tes evaluasi tidak valid. Beberapa faktor tersebut secara garis besar dapat dibedakan menurut sumbernya, yaitu faktor internal dari tes, faktor eksternal tes dan faktor yang berasal dari peserta didik yang bersangkutan.¹⁰

- a) Faktor yang berasal dari dalam tes.

Beberapa sumber yang pada umumnya berasal dari faktor internal tes evaluasi di antaranya sebagai berikut.

- 1) Arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes.
- 2) Kata-kata yang digunakan dalam struktur instrumen evaluasi, terlalu sulit.
- 3) Item-item tes dikonstruksi dengan jelek.
- 4) Tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima peserta didik.
- 5) Waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar.
- 6) Jumlah item tes terlalu sedikit sehingga tidak mewakili sampel materi pembelajaran.

⁹ Zainal Arifin, *Evaluasi Pembelajaran* (Bandung: PT Remaja Rosdakarya, 2011), hlm. 247-248.

¹⁰ Sukardi, *Evaluasi Pendidikan Prinsip dan Operasionalnya*, hlm. 38-39.

- 7) Jawaban masing-masing item evaluasi bisa diprediksi peserta didik.
- b) Faktor yang berasal dari administrasi dan skor. Faktor ini dapat mengurangi validitas interpretasi tes evaluasi, khususnya tes evaluasi yang dibuat oleh guru. Berikut beberapa contoh faktor yang sumbernya berasal dari proses administrasi dan skor.
- 1) Waktu pengerjaan tidak cukup sehingga peserta didik dalam memberikan jawaban dalam situasi yang tergesa-gesa.
 - 2) Adanya kecurangan dalam tes sehingga tidak bisa membedakan antara peserta didik yang belajar dengan yang melakukan kecurangan.
 - 3) Pemberian petunjuk dari pengawas yang tidak dapat dilakukan pada semua peserta didik.
 - 4) Teknik pemberian skor yang tidak konsisten, misalnya pada tes esai, juga dapat mengurangi validitas tes evaluasi.
 - 5) Peserta didik tidak dapat mengikuti arahan yang diberikan dalam tes baku.
 - 6) Adanya joki (orang lain bukan peserta didik) yang masuk dan menjawab item tes yang diberikan.
- c) Faktor-faktor yang berasal dari jawaban peserta didik
- Seringkali terjadi bahwa interpretasi terhadap item-item tes evaluasi tidak valid, karena dipengaruhi oleh jawaban peserta didik daripada interpretasi item-item pada tes evaluasi. Sebagai contoh, sebelum tes para peserta didik menjadi tegang karena guru pengampu mata pelajaran dikenal killer, galak dan sebagainya sehingga peserta didik yang ikut tes banyak yang gagal. Contoh lain, ketika peserta didik melakukan tes penampilan keterampilan, ruangan terlalu ramai atau gaduh sehingga para peserta didik tidak dapat konsentrasi dengan baik. Ini semua dapat mengurangi nilai validitas instrumen evaluasi.

Validitas menurut pendapat beberapa ahli dapat digolongkan dalam beberapa jenis, yakni: validitas konstruksi (*construct validity*), validitas isi (*content validity*), validitas prediktif (*predictive validity*), validitas rupa (*face validity*) dan validitas konkuren (*concurrent validity*). Dalam penelitian ini hanya akan dibahas mengenai validitas isi dan validitas konstruksi.

a) Validitas Isi

Validitas isi (*content validity*) adalah pengujian validitas dilakukan atas isinya untuk memastikan apakah butir tes hasil belajar mengukur secara tepat keadaan yang ingin diukur.¹¹ Validitas isi adalah validitas yang ditilik dari segi isi tes itu sendiri sebagai alat pengukur hasil belajar yaitu: sejauh mana tes hasil belajar sebagai alat pengukur hasil belajar peserta didik, isinya telah dapat mewakili secara representatif terhadap keseluruhan materi atau bahan pelajaran yang seharusnya diteskan (diujikan).¹²

Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan.¹³ Oleh karena materi yang diajarkan tertera dalam kurikulum maka validitas isi ini sering juga disebut validitas kurikuler. Validitas isi dapat diusahakan tercapainya sejak saat penyusunan soal dengan cara merinci materi kurikulum atau materi buku pelajaran.

Untuk menyusun instrumen tes yang mempunyai validitas isi, maka instrumen harus disusun berdasarkan materi pelajaran yang telah dipelajari peserta didik.¹⁴ Untuk mengetahui apakah

¹¹ Purwanto, *Evaluasi Hasil Belajar* (Yogyakarta: Pustaka Pelajar, 2011), hlm. 120.

¹² Anas Sudijono, *Pengantar Evaluasi Pendidikan* (Jakarta: Rajawali Pers, 2009), hlm. 164.

¹³ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 67.

¹⁴ S. Eko Putro Widoyoko, *Evaluasi Program Pembelajaran* (Yogyakarta: Pustaka Pelajar, 2009), hlm. 129-130.

instrumen tes tersebut memiliki validitas isi atau tidak dapat dilakukan dengan membandingkan materi tes tersebut dengan analisis rasional yang dilakukan terhadap bahan-bahan yang seharusnya digunakan dalam penyusunan instrumen tes tersebut. Apabila materi tes tersebut telah cocok dengan analisis rasional yang dilakukan maka instrumen tes yang dianalisis telah memiliki validitas isi. Sebaliknya apabila instrumen tes tersebut tidak cocok dengan analisis rasional yang telah dilakukan maka instrumen tes tersebut tidak memiliki validitas isi.

Pembicaraan tentang validitas isi identik dengan pembicaraan tentang populasi dan sampel. Kalau saja keseluruhan materi pelajaran yang telah diberikan kepada peserta didik atau sudah diperintahkan untuk dipelajari oleh peserta didik kita anggap sebagai populasi, dan isi tes hasil belajar dalam mata pelajaran yang sama kita anggap sebagai sampelnya, maka tes hasil belajar dalam mata pelajaran tersebut dapat dikatakan telah memiliki validitas isi, apabila isi tes tersebut, dapat menjadi wakil yang representatif bagi seluruh materi pelajaran yang telah diajarkan atau telah diperintahkan untuk dipelajari¹⁵. Pengujian validitas isi dapat dilakukan menggunakan satu dari tiga metode yaitu menelaah butir instrumen, meminta pertimbangan ahli dan analisis korelasi butir soal.¹⁶

Dalam pengujian validitas isi bisa dilihat juga dari ranah kognitifnya. Dalam penelitian ini soal akan dianalisis berdasarkan ranah kognitif taksonomi Bloom.

Ranah kognitif dalam Taksonomi Bloom meliputi enam jenjang kemampuan.¹⁷

¹⁵ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 164-165.

¹⁶ Purwanto, *Evaluasi Hasil Belajar*, hlm. 120.

¹⁷ H. Daryanto, *Evaluasi Pendidikan* (Jakarta: Rineka Cipta, 2008), hlm. 103-112.

1) Pengetahuan (*knowledge*)

Pengetahuan adalah aspek yang paling dasar dalam taksonomi Bloom. Seringkali disebut juga aspek ingatan. Dalam jenjang kemampuan ini seseorang dituntut untuk dapat mengenali atau mengetahui adanya konsep, fakta atau istilah-istilah, dan lain sebagainya tanpa harus mengerti atau menggunakannya.

2) Pemahaman (*comprehension*)

Tipe hasil belajar yang lebih tinggi dari pengetahuan adalah pemahaman. Kemampuan ini umumnya mendapat penekanan dalam proses belajar mengajar. Peserta didik dituntut memahami atau mengerti apa yang diajarkan, mengetahui apa yang sedang dikomunikasikan dan dapat memanfaatkan isinya tanpa keharusan menghubungkannya dengan hal-hal lain.

3) Penerapan atau aplikasi (*application*)

Aplikasi adalah penggunaan abstraksi pada situasi konkret atau situasi khusus. Abstraksi tersebut bisa berupa teori, ide atau petunjuk teknis. Menerapkan abstraksi tersebut ke dalam suatu situasi yang baru disebut dengan aplikasi.¹⁸

4) Analisis (*analysis*)

Dalam jenjang kemampuan ini seseorang dituntut untuk dapat menguraiakan suatu situasi atau keadaan tertentu ke dalam unsur-unsur atau komponen-komponen pembentuknya. Dengan jalan ini situasi atau keadaan tersebut menjadi lebih jelas. Bentuk soal yang sesuai untuk mengukur kemampuan ini adalah pilihan ganda dan uraian.

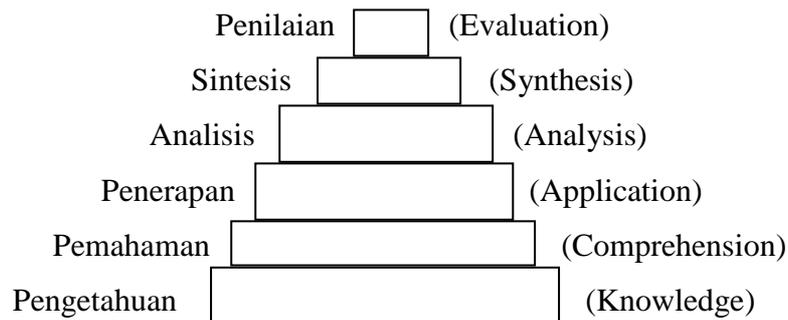
¹⁸ Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar* (Bandung: PT Remaja Rosdakarya, 2010), hlm. 25.

5) Sintesis (*synthesis*)

Pada jenjang ini seseorang dituntut untuk dapat menghasilkan sesuatu yang baru dengan jalan menggabungkan berbagai faktor yang ada.

6) Evaluasi (*evaluation*)

Dalam jenjang kemampuan ini seseorang dituntut untuk dapat mengevaluasi situasi, keadaan, pernyataan, atau konsep berdasarkan suatu kriteria tertentu. Yang penting dalam evaluasi ialah menciptakan kondisinya sedemikian rupa sehingga peserta didik mampu mengembangkan kriteria, standar, atau ukuran untuk mengevaluasi sesuatu. Enam jenjang kemampuan ranah kognitif tersebut dapat digambarkan seperti piramida di bawah ini:



Gambar 1.
Sistem Klasifikasi Bloom

Sedangkan kata kerja operasional yang menunjukkan enam ranah kognitif di atas adalah sebagai berikut:¹⁹

Tabel 1

Kata kerja operasional ranah kognitif pada Taksonomi Bloom

Jenjang kemampuan	Kata kerja operasional
Pengetahuan	Menyebutkan kembali, menghafal, menunjukkan, menggarisbawahi,

¹⁹ Bermawy Munthe, *Desain Pembelajaran* (Yogyakarta: Pustaka Insan Madani, 2010), hlm. 40-42.

	menyortir, menyatakan
Pemahaman	Menjelaskan, mendeskripsikan, membuat pernyataan ulang, menguraikan, menerangkan, mengubah, memberikan contoh, menyadur
Penerapan	Mengoperasikan, mendemonstrasikan, menghitung, menghubungkan, membuktikan, menghasilkan, menunjukan
Analisis	Membandingkan, mempertentangkan, memisahkan, menghubungkan, membuat diagram/skema, menunjukan hubungan, mempertanyakan
Sintesis	Mengategorikan, mengombinasikan, mengarang/menciptakan, mendesain/merancang, menyusun kembali, merangkaikan, menyimpulkan, membuat pola
Penilaian	Mempertahankan, mengategorikan, mengombinasikan, mengarang, menciptakan, mendesain, mengatur, menyusun kembali, merangkaikan, menghubungkan, menyimpulkan, merancang, membuat pola, memberikan argumen

b) Validitas Kontruksi

Secara etimologis, kata konstruksi mengandung arti susunan, kerangka atau rekaan. Dengan demikian, validitas konstruksi dapat diartikan sebagai validitas yang ditilik dari segi susunan, kerangka atau rekaan.²⁰ Borg dan Gall mendefinisikan: *“Construct validity is the extent to which a particular test can be shown to measure hypothetical construct.”*²¹

Secara terminologis, suatu tes hasil belajar dapat dinyatakan sebagai tes yang telah memiliki validitas konstruksi, apabila tes hasil belajar tersebut ditinjau dari segi susunan, kerangka atau rekaannya telah dapat dengan secara tepat

²⁰ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 166.

²¹ Tedjo N. Reksoatmodjo, *Statistika untuk Psikologi dan Pendidikan* (Bandung: PT Refika Aditama, 2009), hlm. 194.

mencerminkan suatu konstruksi dalam teori psikologis.²² Tentang istilah konstruksi dalam teori psikologis ini perlu dijelaskan, bahwa para ahli di bidang psikologis mengemukakan teori yang menyatakan bahwa jiwa dari seseorang peserta didik itu dapat dirinci ke dalam beberapa aspek atau ranah tertentu. Benjamin S. Bloom misalnya merincinya dalam tiga aspek kejiwaan yaitu aspek kognitif, aspek afektif dan aspek psikomotorik.

Validitas konstruksi bukanlah dimaksudkan bahwa tes yang bersangkutan dipandang sudah baik susunan kalimat soalnya, atau urutan butir nomor soalnya sudah runtut, melainkan bahwa tes hasil belajar baru dapat dikatakan telah memiliki validitas konstruksi apabila butir-butir soal atau item yang membangun tes tersebut benar-benar telah dapat dengan secara tepat mengukur aspek-aspek berpikir sebagaimana telah ditentukan dalam tujuan instruksional khusus.

Validitas konstruksi mengacu pada sejauh mana suatu instrumen mengukur konsep dari suatu teori, yaitu yang menjadi dasar penyusunan instrumen. Definisi atau konsep yang diukur berasal dari teori yang digunakan. Oleh karena itu, harus ada pembahasan mengenai teori yang menjadi dasar penentuan konstruksi suatu instrumen.

2) Reliabilitas

Keandalan (*reliability*) berasal dari kata *rely* yang artinya percaya dan reliabel yang artinya dapat dipercaya. Keterpercayaan berhubungan dengan ketepatan dan konsistensi.²³ Reliabilitas berarti konsistensi di mana suatu instrumen menghasilkan hasil skor yang

²² Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 166.

²³ Purwanto, *Evaluasi Hasil Belajar*, hlm. 153-154.

sama.²⁴ Reliabilitas alat ukur (instrumen) adalah ketetapan atau keajekan instrumen tersebut dalam mengukur apa yang diukurnya.²⁵ Reliabilitas mempunyai berbagai nama lain seperti keterpercayaan, keterandalan, keajegan, kestabilan, konsistensi dan sebagainya, namun ide pokok yang terkandung dalam konsep reliabilitas adalah sejauh mana suatu pengukuran dapat dipercaya.²⁶ Berikut adalah pendapat para ahli mengenai reliabilitas:

- a) Menurut Thorndike dan Hagen, reliabilitas berhubungan dengan akurasi instrumen dalam mengukur apa yang diukur, kecermatan hasil ukur dan seberapa akurat seandainya dilakukan pengukuran ulang.
- b) Hopkins dan Antes menyatakan reliabilitas sebagai konsistensi pengamatan yang diperoleh dari pencatatan berulang baik pada satu subjek maupun sejumlah subjek.²⁷
- c) Conny Semiawan mengungkapkan bahwa pengertian reliabilitas menunjuk pada ketetapan (konsistensi) dari nilai yang diperoleh sekelompok individu dalam kesempatan yang berbeda dengan tes yang sama ataupun yang itemnya ekuivalen.²⁸

Dari beberapa definisi reliabilitas tersebut dapat disimpulkan bahwa reliabilitas berhubungan dengan kemampuan alat ukur untuk melakukan pengukuran secara cermat. Reliabilitas merupakan akurasi dan presisi yang dihasilkan oleh alat ukur dalam melakukan pengukuran. Alat ukur yang reliabel akan menghasilkan ukuran yang

²⁴ M. Djunaidi Ghoni dan Fauzan Almanshur, *Petunjuk Praktis Penelitian Pendidikan* (Malang: UIN-Malang Press, 2009), hlm. 234.

²⁵ Mawardi Lubis, *Evaluasi Pendidikan Nilai* (Yogyakarta: Pustaka Pelajar, 2009), hlm. 45.

²⁶ Saifuddin Azwar, *Reliabilitas dan Validitas* (Yogyakarta: Pustaka Pelajar, 2011), hlm. 4.

²⁷ Purwanto, *Evaluasi Hasil Belajar*, hlm. 154.

²⁸ Dewa Ketut Sukardi dan Nila kusmawati, *Analisis Tes Psikologis Teori & Praktik* (Jakarta: Rineka Cipta, 2009), hlm. 259.

sebenarnya. Alat ukur yang reliabel akan memberikan hasil pengukuran yang relatif stabil dan konsisten karena pengukurannya menghasilkan galat yang minimal.

Setiap alat pengukur seharusnya memiliki kemampuan untuk memberikan hasil pengukuran yang konsisten. Makin kecil kesalahan yang terjadi dalam pengukuran, maka semakin reliabel alat pengukur tersebut. Sebaliknya semakin besar kesalahan pengukuran maka semakin tidak reliabel alat pengukur tersebut.

Reliabilitas suatu tes pada umumnya dituliskan secara numerik dalam bentuk koefisien. Apabila koefisiennya tinggi maka menunjukkan reliabilitas yang tinggi. Sebaliknya jika koefisiennya rendah maka reliabilitas tesnya rendah juga. Tidak reliabelnya suatu tes hasil belajar pada prinsipnya dapat dikatakan bahwa tes hasil belajar tersebut sia-sia, hal ini dikarenakan jika dilakukan pengesanan kembali maka hasilnya akan berbeda.

Ada beberapa faktor yang dapat mempengaruhi reliabilitas (keandalan), diantaranya adalah:²⁹

a) Luas tidaknya sampling yang diambil

Dengan semakin luasnya suatu sampling, maka tes hasil belajar akan semakin reliabel

b) Perbedaan bakat dan kemampuan peserta didik yang dites

Tes yang diberikan terhadap peserta didik dengan bakat dan kemampuan yang berbeda maka akan menghasilkan reliabilitas yang berbeda pula.

c) Suasana dan kondisi saat berlangsungnya tes.

Suasana dan keadaan saat berlangsungnya tes, seperti gaduh, tenang, banyak gangguan dapat mempengaruhi hasil dan reliabilitas dari tes tersebut.

²⁹ M. Ngalim Purwanto, *Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran* (Bandung: PT Remaja Rosdakarya, 2009), hlm. 141.

Ada beberapa cara untuk mengetahui reliabilitas suatu tes, diantaranya adalah:

a) Metode bentuk paralel (*equivalent*)

Pendekatan reliabilitas bentuk paralel dilakukan dengan memberikan sekaligus dua bentuk tes yang paralel satu sama lain, kepada sekelompok subjek.³⁰ Sesuai dengan namanya, yaitu ekuivalen maka tes yang hendak diukur reliabilitasnya dibuat identik. Setiap tampilannya, kecuali sunstansi item yang ada dapat berbeda. Kedua tes tersebut sebaiknya mempunyai karakteristik sama. Karakteristik yang dimaksud termasuk, misalnya: mengukur variabel yang sama, mempunyai jumlah item sama, struktur sama, mempunyai tingkat kesulitan sama dan mempunyai petunjuk, cara skoring dan interpretasi yang sama.³¹

Berikut ini adalah langkah-langkah melaksanakan tes reliabilitas secara ekuivalen:³²

- 1) Tentukan subjek sasaran yang hendak dites.
- 2) Lakukan tes yang dimaksud kepada subjek sasaran tersebut.
- 3) Administrasikan hasilnya secara baik.
- 4) Dalam waktu yang tidak terlalu lama, lakukan pengetesan untuk yang kedua kalinya pada grup tersebut.
- 5) Korelasikan kedua hasil tes skor.

Menentukan korelasi antara tes pertama dan kedua menggunakan rumus korelasi product moment sebagai berikut:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

³⁰ Saifuddin Azwar, *Reliabilitas dan Validitas*, hlm. 59.

³¹ Sukardi, *Metodologi Penelitian Pendidikan* (Jakarta: PT Bumi Aksara, 2011), hlm. 129.

³² Sukardi, *Metodologi Penelitian Pendidikan*, hlm. 130.

b) Metode tes ulang (*test-retest method*)

Metode yang jelas untuk mengetahui reliabilitas suatu tes adalah dengan cara memberikan tes itu dua kali kepada sekelompok individu yang sama.³³ Metode pengujian reliabilitas yang dilakukan dengan mengujikan sebuah perangkat tes hasil belajar kepada kelompok peserta uji coba yang sama sebanyak dua kali disebut dengan metode tes ulang. Hasil pengukuran kedua pengujian selanjutnya dikorelasikan.

Instrumen penelitian yang reliabilitasnya diuji dengan metode tes ulang dilakukan dengan cara mencobakan instrumen dua kali atau lebih kepada responden. Dalam hal ini, instrumennya sama, respondennya sama tetapi waktunya yang berbeda.³⁴ Sebuah tes hasil belajar dikatakan reliabel apabila dua kali atau lebih pengujian menunjukkan hasil yang stabil. Stabilitas ditunjukkan oleh korelasi antara skor yang diperoleh dari kedua pengujian.³⁵ Pelaksanaan metode ini adalah sebagai berikut:³⁶

- 1) Pada bulan November disajikan suatu bentuk tes misalnya tes matematika kepada peserta didik kelas X SMA.
- 2) Setelah beberapa waktu setelah tes pertama, misalnya saja pada bulan Desember tes matematika tersebut diteskan kembali kepada peserta didik kelas X SMA.
- 3) Skor perolehan dari kedua tes tersebut kemudian dikorelasikan untuk mengestimasi reliabilitas tes.

Menentukan korelasi antara tes pertama dan kedua menggunakan rumus korelasi product moment sebagai berikut:

³³ H. Arief Furchan, *Pengantar Penelitian Dalam Pendidikan* (Yogyakarta: Pustaka Pelajar, 2007), hlm. 318.

³⁴ Sugiyono, *Statistika Untuk Penelitian* (Bandung: CV Alfabeta, 2008), hlm. 354.

³⁵ Purwanto, *Evaluasi Hasil Belajar*, hlm. 156.

³⁶ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004* (Bandung: PT Remaja Rosdakarya, 2006), hlm. 93.

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Jarak atau selang waktu antara tes pertama dengan tes kedua sebaiknya tidak terlalu dekat dan juga tidak terlalu jauh. Jika terlalu dekat, hasilnya banyak dipengaruhi oleh ingatan peserta didik tentang jawaban yang diberikan pada saat pengukuran pertama. Sebaliknya, jika selang waktu terlalu lama bisa terjadi adanya suatu perubahan pengetahuan dan pengalaman peserta tes sehingga hal ini dapat mempengaruhi reliabilitasnya.

c) Metode belah dua (*split-half method*)

Reliabilitas belah dua ini, termasuk reliabilitas yang mengukur konsistensi internal. Yang dimaksud konsistensi internal adalah salah satu tipe reliabilitas yang didasarkan pada keajegan dalam tes. Reliabilitas belah dua ini pelaksanaannya hanya memerlukan waktu satu kali.³⁷

Ada dua cara membelah butir soal, yaitu:³⁸

- 1) Membelah atas item-item genap dan item-item ganjil yang selanjutnya belahan ganjil-genap.
- 2) Membelah atas item-item awal dan item-item akhir yaitu separuh jumlah pada nomor-nomor awal dan separuh pada nomor-nomor akhir yang selanjutnya disebut belahan awal-akhir.

Dengan menggunakan metode belah dua dapat diperoleh ukuran reliabilitas dari sekali pemberian satu bentuk tes. Tes tersebut diberikan kepada sekelompok subjek, kemudian butir-butir tes tersebut dibagi menjadi dua bagian yang sebanding. Setelah itu dicari skor tiap-tiap individu di kedua bagian itu dan dihitung koefisien korelasi kedua skor tersebut. Apabila setiap subjek mempunyai kedudukan yang serupa di kedua bagian tes tersebut,

³⁷ Sukardi, *Metodologi Penelitian Pendidikan*, hlm. 130.

³⁸ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 93.

maka tes tersebut mempunyai reliabilitas yang tinggi. Jika konsistensi kedudukan tersebut kecil, maka reliabilitas tersebut rendah.

Dalam metode belah dua ini ada beberapa formula untuk menghitung koefisien reliabilitas suatu tes, di antaranya adalah:

1) Formula Spearman-Brown

Langkah-langkah yang perlu ditempuh dalam penentuan reliabilitas dengan menggunakan formula Spearman-Brown adalah sebagai berikut:

- a) Menjumlahkan skor-skor dari butir-butir soal pada belahan pertama yang dimiliki oleh masing-masing peserta tes.
- b) Menjumlahkan skor-skor dari butir-butir soal pada belahan kedua yang dimiliki oleh masing-masing peserta tes.
- c) Menghitung koefisien korelasi “r” product moment dengan rumus:

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

- d) Menghitung koefisien reliabilitas tes dengan menggunakan rumus:

$$r_{11} = \frac{2r_{xy}}{1 + r_{xy}}$$

- e) Memberikan interpretasi terhadap r_{11} .

2) Formula Flanagan

Persamaan lain yang dapat digunakan untuk menentukan reliabilitas belah dua adalah persamaan Flanagan yaitu:³⁹

$$r_{11} = 2 \left(1 - \frac{S_1^2 + S_2^2}{S_t^2} \right)$$

r_{11} = reliabilitas tes

S_1^2 = varian belahan pertama

³⁹ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*, hlm. 107.

S_2^2 = varian belahan kedua

S_t^2 = varian total

3) Formula Rulon

Rulon merumuskan suatu formula untuk mengestimasi reliabilitas belah dua tanpa perlu berasumsi bahwa kedua belahan mempunyai varian yang sama. Formula Rulon dirumuskan sebagai berikut:⁴⁰

$$r_{xx'} = 1 - \frac{S_d^2}{S_x^2}$$

Di mana:

$r_{xx'}$ = reliabilitas tes

S_d^2 = varian perbedaan skor kedua belahan

S_x^2 = varian skor tes

d = perbedaan skor kedua belahan

d) Internal Konsistensi

Metode lainnya yang digunakan untuk menentukan reliabilitas adalah internal konsistensi yang berkaitan dengan unsur-unsur yang membentuk sebuah tes, yaitu soal-soal yang membentuk tes. Terdapat beberapa teknik dan persamaan yang digunakan untuk mencari reliabilitas dengan internal konsistensi yaitu:

1) Koefisien Alpha

Koefisien alpha dihitung dengan menggunakan persamaan berikut:⁴¹

$$\alpha = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum S_j^2}{S_x^2} \right)$$

Di mana:

k = banyaknya belahan tes

⁴⁰ Saifuddin Azwar, *Reliabilitas dan Validitas*, hlm. 72.

⁴¹ Saifuddin Azwar, *Reliabilitas dan Validitas*, hlm. 78.

s_j^2 = varian belahan j ; $j = 1, 2, \dots, k$

s_x^2 = varian skor tes

2) Kuder-Richardson

Adapun formula yang diajukan oleh Kuder-Richardson ada dua buah yang masing-masing diberi kode KR_{20} dan KR_{21} , yaitu:

Rumus KR_{20} ⁴²:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \sum pq}{S^2} \right)$$

Di mana:

r_{11} = reliabilitas tes secara keseluruhan

S^2 = varian total

p = proporsi subyek yang menjawab benar pada suatu butir

q = proporsi subyek yang menjawab item salah ($q = 1-p$)

n = banyaknya item

$\sum pq$ = jumlah hasil kali antara p dan q

Rumus KR_{21} ⁴³:

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

Di mana:

r_{11} = koefisien reliabilitas tes

n = banyaknya butir soal

1 = bilangan konstan

M_t = rata-rata hitung dari skor total

S_t^2 = varian total

3) Teknik Hoyt

Perhitungan koefisien reliabilitas menggunakan teknik Hoyt dilakukan dengan rumus berikut:⁴⁴

⁴² Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 100.

⁴³ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 253.

⁴⁴ Purwanto, *Evaluasi Hasil Belajar*, hlm. 172.

$$r_{11} = 1 - \frac{V(s)}{V(r)}$$

Di mana:

r_{11} = koefisien reliabilitas

$V(s)$ = varian sisa

$V(r)$ = varian responden

b. Analisis Butir Soal Tes Hasil Belajar

Analisis butir soal adalah suatu prosedur yang sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir tes yang kita susun. Kegiatan menganalisis butir soal merupakan suatu kegiatan yang dilakukan untuk meningkatkan mutu soal yang telah ditulis. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban peserta didik untuk membuat keputusan tentang setiap penilaian. Faedah mengadakan analisis butir soal adalah sebagai berikut:⁴⁵

- a. Membantu kita dalam mengidentifikasi butir-butir soal yang jelek.
- b. Memperoleh informasi yang akan dapat digunakan untuk menyempurnakan soal-soal untuk kepentingan lebih lanjut.
- c. Memperoleh gambaran secara selintas tentang keadaan soal yang disusun.

1) Tingkat Kesukaran

Tingkat kesukaran (*difficulty index*) dapat didefinisikan sebagai proporsi peserta didik peserta tes yang menjawab benar.⁴⁶ Soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang peserta didik untuk mempertinggi usaha memecahkannya. Sebaliknya soal yang terlalu sukar akan menyebabkan

⁴⁵ H. Daryanto, *Evaluasi Pendidikan*, hlm. 177-178.

⁴⁶ Purwanto, *Evaluasi Hasil Belajar*, hlm. 99.

peserta didik menjadi putus asa dan tidak mempunyai semangat untuk mencoba lagi karena di luar jangkauannya.

Dalam penyusunan instrumen tes perlu adanya keseimbangan dari tingkat kesukaran tes tersebut. Keseimbangan di sini adalah adanya soal-soal yang termasuk dalam kategori mudah, sedang dan sukar secara proporsional. Hal penting yang perlu diperhatikan dalam melakukan analisis tingkat kesukaran soal yakni penentuan proporsi dan kriteria soal yang termasuk dalam kategori mudah, sedang dan sukar.

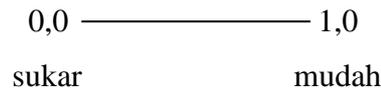
Perbandingan proporsi antara soal yang mudah, sedang dan sukar misalnya bisa dibuat 3-4-3. Ini berarti 30% soal kategori mudah, 40% soal kategori sedang dan 30% soal kategori sukar. Perbandingan lain misalnya, 3-5-2. Ini berarti tes tersebut terdiri dari 30% soal dengan kategori mudah, 50% kategori sedang dan sisanya 20% termasuk dalam kategori sukar.

Dalam menentukan kriteria soal (mudah, sedang, sukar) menggunakan keputusan dari pembuat soal berdasarkan pertimbangan-pertimbangan tertentu, antara lain⁴⁷:

- a. Abilitas yang hendak diukur dalam pertanyaan tersebut. Misalnya dalam aspek kognitif, pengetahuan dan pemahaman termasuk dalam kategori mudah, penerapan dan analisis termasuk dalam kategori sedang dan sintesis dan evaluasi termasuk dalam kategori sukar.
- b. Sifat materi yang diujikan atau ditanyakan. Misalnya, fakta termasuk kategori mudah, konsep dan prinsip termasuk kategori sedang dan generalisasi termasuk kategori sukar.
- c. Isi bahan yang ditanyakan sesuai dengan bidang keilmuannya.
- d. Bentuk soal.

Bilangan yang menunjukkan sukar dan mudahnya sesuatu soal disebut indeks kesukaran. Besarnya indeks kesukaran antara 0,00 sampai dengan 1,00. Indeks kesukaran ini menunjukkan tingkat kesukaran soal. Soal dengan indeks kesukaran 0,00 menunjukkan bahwa soal itu terlalu sukar, sebaliknya indeks 1,00 menunjukkan bahwa soalnya terlalu mudah.

⁴⁷ Nana Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, hlm. 136.



Secara umum tingkat kesukaran dapat dinyatakan melalui beberapa cara diantaranya proporsi menjawab benar, skala kesukaran linier dan indeks Davis.

a. Proporsi menjawab benar

Persamaan yang digunakan untuk menentukan tingkat kesukaran dengan proporsi menjawab benar adalah:⁴⁸

$$p = \frac{\sum x}{S_m N}$$

p = proporsi menjawab benar atau tingkat kesukaran

$\sum x$ = banyaknya peserta tes yang menjawab benar

S_m = skor maksimum

N = jumlah peserta tes

b. Skala kesukaran linier

Skala kesukaran linier ini disusun dengan cara mentransformasikan nilai p menjadi nilai z , di mana perubahan nilai p menjadi nilai z itu dilakukan dengan berkonsultasi pada tabel nilai z . Langkah-langkah menentukan tingkat kesukaran dengan skala kesukaran linier adalah sebagai berikut:

1) Mengoreksi nilai p kotor (p_k) menjadi nilai p bersih (p_b) dengan menggunakan rumus:⁴⁹

$$p_b = \frac{ap_k - 1}{a - 1}$$

p_b = p bersih

p_k = p kotor

a = alternatif atau option yang disediakan atau dipasangkan pada butir item yang bersangkutan.

⁴⁸ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas dan Interpretasi Hasil Tes Implementasi Kurikulum 2004*, hlm. 12.

⁴⁹ Anas Sudijono, *Pengantar Evaluasi Pendidikan*, hlm. 380.

1 = bilangan konstan

2) Mentransformasikan nilai p_b menjadi nilai z dengan berkonsultasi pada tabel kurva normal.

c. Indeks Davis

Dengan menggunakan indeks Davis maka derajat kesukaran item dibuat bergerak antara 0 sampai dengan 100. Dalam keadaan seperti ini tidak akan ada tanda minus.

Langkah-langkah menentukan tingkat kesukaran dengan indeks Davis adalah:

1) Mentransformasikan nilai p menjadi nilai z dengan berkonsultasi pada tabel kurva normal.

2) Menghitung indeks Davis dengan rumus:

$$D = 21,063 z + 50$$

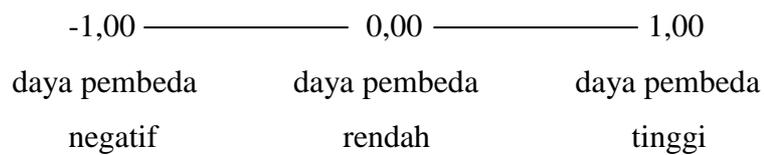
2) Daya Pembeda

Daya pembeda adalah kemampuan butir soal tes hasil belajar membedakan peserta didik yang mempunyai kemampuan tinggi dan rendah. Daya pembeda berhubungan dengan derajat kemampuan butir membedakan dengan baik perilaku pengambil tes dalam tes yang dikembangkan. Daya pembeda harus diusahakan positif dan setinggi mungkin. Butir soal yang mempunyai daya pembeda positif dan tinggi berarti butir tersebut dapat membedakan dengan baik peserta didik yang berkemampuan tinggi dan yang berkemampuan rendah.⁵⁰

Mengetahui daya pembeda item itu penting sekali, sebab salah satu dasar yang dipegangi untuk menyusun butir-butir item tes hasil belajar adalah adanya anggapan, bahwa kemampuan antara peserta didik yang satu dengan peserta didik yang lain itu berbeda-beda, dan bahwa butir-butir item tes hasil belajar itu haruslah mampu memberikan hasil tes yang mencerminkan adanya perbedaan-perbedaan kemampuan yang terdapat di kalangan peserta didik tersebut.

⁵⁰ Purwanto, *Evaluasi Hasil Belajar*, hlm. 102.

Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi. Seperti halnya indeks kesukaran, indeks diskriminasi ini berkisar antara 0,00 sampai 1,00. Hanya bedanya, indeks kesukaran tidak mengenal tanda negatif (-), tetapi pada indeks diskriminasi ada tanda negatif. Tanda negatif pada indeks diskriminasi digunakan jika suatu soal terbalik menunjukkan kualitas testee. Yaitu anak pandai disebut bodoh dan anak bodoh disebut pandai.



Untuk mengetahui besar kecilnya angka indeks diskriminasi item dapat dipergunakan dua macam rumus sebagai berikut:

Rumus pertama:

$$D = \frac{\sum A}{n_A} - \frac{\sum B}{n_B}$$

Di mana:

D = indeks diskriminasi

$\sum A$ = jumlah peserta tes yang menjawab benar pada kelompok atas

$\sum B$ = jumlah peserta tes yang menjawab benar pada kelompok bawah

n_A = jumlah peserta tes kelompok atas

n_B = jumlah peserta tes kelompok bawah

Rumus kedua:

Angka indeks diskriminasi item diperoleh dengan menggunakan teknik korelasi Phi dengan rumus:

$$\phi = \frac{P_H - P_L}{2\sqrt{(p)(q)}}$$

Di mana:

ϕ = angka indeks korelasi phi, yang dalam hal ini dianggap sebagai angka indeks diskriminasi item

P_H = proporsi peserta tes yang menjawab benar pada kelompok atas

P_L = proporsi peserta tes yang menjawab benar pada kelompok bawah

2 = bilangan konstan

p = proporsi seluruh peserta tes yang menjawab benar

q = proporsi seluruh peserta tes yang menjawab salah

Bagi suatu soal yang dapat dijawab benar oleh peserta didik pandai maupun bodoh, maka soal itu tidak baik karena tidak mempunyai daya pembeda. Demikian pula jika semua peserta didik baik pandai maupun bodoh tidak dapat menjawab dengan benar. Soal yang baik adalah soal yang dapat dijawab benar oleh peserta didik yang pandai saja.⁵¹

Dalam menghitung daya pembeda terdapat beberapa kejadian khusus yang harus diperhatikan:⁵²

- a. Bila data di tengah sama maka data yang sama dikeluarkan dari analisis.
- b. Dalam hal jumlah peserta didik uji coba sangat banyak maka penentuan kelompok atas dan bawah adalah dengan mengambil 27% peserta didik yang memperoleh skor tertinggi sebagai kelompok atas dan 27% peserta didik yang memperoleh skor terendah sebagai kelompok bawah. Sebanyak 46% peserta didik di tengah distribusi dikeluarkan dan tidak dianalisis.

⁵¹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, hlm. 211.

⁵² Purwanto, *Evaluasi Hasil Belajar*, hlm. 107.