

MENYUSUN DAN MENGANALISIS TES HASIL BELAJAR

Abdul Kadir

Fakultas Tarbiyah dan Ilmu Keguruan IAIN Kendari

Email : abdir_edu@yahoo.co.id

Abstrak

Tes merupakan suatu bentuk alat evaluasi untuk mengukur seberapa jauh tujuan pengajaran telah tercapai, jadi berarti evaluasi terhadap hasil belajar. Tes yang baik harus memenuhi beberapa persyaratan, yaitu; harus efisien, harus baku, mempunyai norma, objektif, valid (*sahih*), dan reliabel (*andal*). Untuk memperoleh tes yang memenuhi persyaratan tersebut maka tes yang telah dibuat perlu dianalisis. Analisis tes dimulai dari saat menyusun tes dimana tes yang disusun harus berdasarkan Silabus/SAP setiap mata pelajaran, membuat kisi-kisi terlebih dahulu, baru kemudian menyusun soal sesuai kaidah-kaidah penyusunan soal berdasarkan jenis soal yang diinginkan. Bentuk tes pilihan ganda sukar atau mudahnya suatu soal bukan semata-mata ditentukan oleh materi soal, akan tetapi ditentukan juga oleh teknik penyusunannya. Pedoman umum penulisan butir soal tes pilihan ganda adalah; 1) Butir soal harus sesuai dengan indikator, 2) Pokok soal dan pilihan jawaban harus dirumuskan secara jelas, singkat, padat, dan tegas, sehingga perumusan tersebut hanya mencakup pernyataan yang diperlukan saja, 3) Pokok soal jangan memberi petunjuk ke arah jawaban yang benar, 4) Pokok soal dan pilihan jawaban tidak mengandung pernyataan yang bersifat negatif ganda, 5) Pilihan jawaban yang merupakan kunci jawaban harus menunjukkan kebenaran mutlak dan terbaik, 6) Pilihan jawaban harus homogen dan logis secara materi dan bahasa, 7) Panjang rumusan pilihan jawaban harus relatif sama, 8) Pilihan jawaban sebaiknya jangan memakai bunyi "semua pilihan jawaban di atas salah" atau "semua pilihan jawaban di atas benar", dan 9) Pilihan jawaban berbentuk angka harus disusun berdasarkan urutan kecil ke besar atau sebaliknya. Tes yang telah disusun perlu diuji coba terlebih dahulu kepada peserta lain sebelum diujikan kepada peserta yang sebenarnya. Selanjutnya setiap peserta uji coba diambil skornya masing-masing dan ditabelkan, kemudian hitung validitasnya, tingkat kesukaran, dan daya beda. Sebagai kesimpulan dari analisis tes adalah bahwa dari sejumlah tes yang disusun maka soal yang diambil adalah soal-soal yang valid, indek kesukaran sedang, daya beda baik dan baik sekali. Soal yang tidak memenuhi kriteria ini dibuang (tidak dipakai). Oleh sebab itu jumlah soal yang disusun harus melebihi jumlah soal yang diinginkan.

Kata Kunci: validitas, reliabilitas, tes, hasil belajar

A. PENDAHULUAN

Upaya untuk mengukur seberapa jauh tujuan-tujuan pembelajaran yang telah tercapai, dapat dilakukan dengan evaluasi, dalam hal ini evaluasi hasil belajar. Alat ukur untuk mengevaluasi hasil belajar tersebut digunakan tes. Tes adalah cara (yang dapat dipergunakan) atau prosedur yang (yang perlu ditempuh) dalam rangka pengukuran dan penilaian di bidang pendidikan. Sebagai pendidik, keterampilan yang harus dikuasai adalah sistem penilaian hasil belajar peserta didik. Dalam penilaian proses dan hasil belajar siswa di sekolah, aspek yang berkenaan dengan pemilihan alat penilaian, yaitu; penyusunan soal, analisis butir soal untuk memperoleh kualitas soal yang memadai, dan pengolahan dan interpretasi data hasil penilaian. Menurut Arikunto, tes yang baik harus mempunyai syarat-syarat antara lain: 1) harus efisien (*parsimony*), 2) harus baku (*standardize*), 3) mempunyai norma, 4) objektif, 5) valid (sahih), dan 6) reliabel (andal).¹ Oleh sebab itu untuk memperoleh tes yang baik, tes tersebut harus diujicobakan terlebih dahulu dan hasilnya dianalisis sehingga memenuhi syarat-syarat tersebut di atas.

Salah satu bentuk tes hasil belajar adalah tes pilihan ganda. Tes pilihan ganda adalah bentuk tes obyektif yang mempunyai ciri utama kunci jawaban jelas dan pasti sehingga hasilnya dapat diskor secara obyektif. Artinya setelah siswa mengerjakan soal dalam bentuk tes pilihan ganda maka siswa tersebut akan memperoleh skor yang sama jika hasil pekerjaannya diperiksa oleh lebih dari satu pemeriksa. Hal ini disebabkan setiap jawaban diberi skor yang sudah pasti dan tidak mengenal jawaban di antara benar dan salah atau jawaban benar sebagian saja. Soal pilihan ganda terdiri dari pernyataan dan pertanyaan yang harus dijawab oleh siswa atau melengkapi dengan memilih salah satu dari beberapa alternatif yang tersedia. Satu di antaranya adalah yang paling benar, lainnya disebut pengecoh (*distractor*).

Noeng Muhajir mengatakan bahwa pengertian tes pilihan ganda merupakan tes objektif dimana masing-masing item disediakan lebih dari dua kemungkinan jawaban, dan hanya satu dari pilihan-pilihan tersebut yang benar atau yang paling benar.² Sedangkan keunggulan tes pilihan ganda menurut Azwar, yaitu; 1) kompherhensif, karena dalam waktu tes yang singkat dapat memuat lebih banyak item, 2) pemeriksaan jawaban dan pemberian skornya mudah dan cepat, 3) penggunaan lembar jawaban

¹ Arikunto Suharsimi, *Dasar-Dasar Evaluasi Pendidikan*, (Jakarta: Bumi Aksara, 2003), h.153

² Chabib Moh. Thoha. *Teknik Evaluasi Pendidikan*, (Jakarta: PT Raja Grafindo Persada, 2001)

menjadikan tes efisien dan hemat bahan, 4) kualitas item dapat dianalisis secara empiric, 5) objektivitasnya tinggi, dan 6) umumnya memiliki reabilitas yang memuaskan.³ Disamping keunggulan tes pilihan ganda mempunyai kelemahan yaitu; 1) pembuatannya sulit dan memakan banyak waktu dan tenaga, 2) tidak mudah ditulis untuk mengungkapkan tingkat kompetensi tinggi, 3) ada kemungkinan jawaban benar semata-mata karena tebakan.

Dalam tulisan ini, penulis mencoba menguraikan bagaimana cara menyusun dan menganalisis tes. Analisis tes dapat dilakukan dengan jalan menentukan/menghitung: 1) Validitas butir tes (*r-butir*), 2) Indeks kesukaran, 3) Daya pembeda, 4) Pengecoh, dan 5) Reliabilitas tes. Tujuan melakukan analisis tes adalah untuk mendapatkan tes yang baik sesuai persyaratan kelayakan instrument penilaian.

B. LANGKAH-LANGKAH MENYUSUN SOAL

1. Merujuk pada Silabus/SAP

Biasanya suatu sekolah/lembaga pendidikan telah mempunyai Silabus dan SAP untuk setiap mata pelajaran. Silabus berisikan pokok-pokok bahasan yang akan diajarkan dalam satu semester. Silabus diperlukan pada waktu membuat kisi-kisi soal agar soal yang dibuat mewakili semua pokok bahasan yang ada sehingga akhirnya dapat dilihat apakah tujuan pembelajaran tercapai atau tidak.

2. Menyusun Kisi-Kisi Soal

Menyusun kisi-kisi merupakan langkah awal yang harus dilakukan setiap kali menyusun tes dan menulis soal. Dengan adanya kisi-kisi, penyusunan soal dapat menghasilkan tes yang relatif sama. Kisi-kisi tes adalah suatu format atau matriks yang memuat kriteria butir soal yang diperlukan dalam menyusun tes. Oleh karena itu, kisi-kisi yang baik harus memenuhi beberapa kriteria, yaitu; 1) dapat menggambarkan keterwakilan isi kurikulum, 2) komponen yang membentuk kisi-kisi harus jelas, rinci, dan mudah dipahami, dan 3) Setiap indikator dapat dituliskan butir soalnya.

3. Menyusun Soal

Soal dapat disusun dalam bentuk tes objektif maupun tes esai. Sebagai bahasan dalam tulisan ini penulis memilih bentuk tes objektif dengan bentuk soal tes pilihan ganda. Jumlah soal yang disusun harus melebihi jumlah yang dibutuhkan dan disusun sesuai kisi-kisi. Sukar atau mudahnya suatu soal bukan semata-mata ditentukan oleh materi soal, akan tetapi ditentukan juga oleh teknik penyusunannya. Beberapa butir

³ Syarifuddin, Azwar. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*, (edisi II, cetakan ke 4 : Pustaka Pelajar, 2003)

pernyataan yang merupakan bagian pokok dalam pedoman umum penulisan butir soal tes pilihan ganda adalah sebagai berikut:

- a. Butir soal harus sesuai dengan indikator,
- b. Pokok soal dan pilihan jawaban harus dirumuskan secara jelas, singkat, padat, dan tegas, sehingga perumusan tersebut hanya mencakup pernyataan yang diperlukan saja,
- c. Pokok soal jangan memberi petunjuk ke arah jawaban yang benar,
- d. Pokok soal dan pilihan jawaban tidak mengandung pernyataan yang bersifat negatif ganda,
- e. Pilihan jawaban yang merupakan kunci jawaban harus menunjukkan kebenaran mutlak dan terbaik,
- f. Pilihan jawaban harus homogen dan logis secara materi dan bahasa,
- g. Panjang rumusan pilihan jawaban harus relatif sama,
- h. Pilihan jawaban sebaiknya jangan memakai bunyi “semua pilihan jawaban di atas salah” atau “semua pilihan jawaban di atas benar”, dan
- i. Pilihan jawaban berbentuk angka harus disusun berdasarkan urutan kecil ke besar atau sebaliknya.

4. Melaksanakan Uji Coba Tes

Agar memperoleh soal/tes yang baik maka soal/test tersebut harus diuji coba terlebih dahulu dan hasilnya dianalisis sehingga memenuhi syarat-syarat tes yang baik. Peserta uji coba misalnya adalah siswa, maka siswatersebut harus mempunyai status sama dengan peserta tes yang sebenarnya.

5. Membuat Skor

Setelah soal diuji coba maka selanjutnya dibuat skor masing-masing siswa (peserta yang diuji coba). Bila siswa menjawab benar diberi skor 1, dan bila siswa menjawab salah atau tidak menjawab diberi skor 0. Semua skor yang diperoleh untuk setiap siswa dibuat dalam bentuk tabel.

C. LANGKAH-LANGKAH MENGANALISIS TES

1. Menghitung Validitas Butir (*r butir*)

Setiap item soal dianalisis validitas butirnya dengan cara sebagai berikut:

- a. Tabel skor diurutkan dari total skor tertinggi ke terendah,
- b. Setiap butir soal dihitung nilai validitasnya dengan menggunakan rumus *korelasi product moment pearson*:

$$r_{xy} = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{n\Sigma X^2 - (\Sigma X)^2\}\{n\Sigma Y^2 - (\Sigma Y)^2\}}} \dots^4$$

Keterangan:

- r_{xy} = Koefisien korelasi antara variabel X dengan variabel Y.
 N = Jumlah Responden.
 ΣX = Jumlah skor item variabel X
 ΣY = Jumlah skor variabel Y
 ΣXY = Jumlah skor dalam sebaran X dan Y
 ΣX^2 = Jumlah kuadrat dari jumlah skor dalam sebaran X
 ΣY^2 = Jumlah kuadrat dari jumlah skor dalam sebaran Y

- c. Harga/nilai koefisien korelasi (r) dikonfirmasi dengan tabel kritik *product moment* pada taraf signifikansi 5% ($\alpha = 0,05$) dan dk $n-1$. Kriteria nilai koefisien korelasi tersebut dikategorikan sebagai berikut:
- $0,8 \leq r \leq 1$ sangat tinggi
 - $0,6 \leq r \leq 0,79$ tinggi
 - $0,4 \leq r \leq 0,59$ cukup
 - $0,2 \leq r \leq 0,39$ rendah
 - $0,0 \leq r \leq 0,19$ rendah sekali.⁵

Apabila hasil perhitungan validitas butir untuk soal tertentu diperoleh r -hitung lebih besar jika dibandingkan dengan r -tabel pada taraf signifikansi 5% ($\alpha = 0,05$) dan dk $n-1$ berarti soal tersebut valid. Perhitungan serupa dilakukan untuk semua soal yang ada. Terkait dengan penafsiran nilai koefisien korelasi r , Widoyoko dan Sugiyono mengatakan, bahwa harga kritik untuk validitas instrumen adalah 0,3. Artinya, bahwa butir soal dikatakan valid jika nilai koefisien korelasi 0,3 atau lebih (paling kecil 0,3) dan sebaliknya dikategorikan butir soal tidak valid.⁶Dari sejumlah soal yang divalidasi, maka kemungkinan terdapat beberapa soal yang tidak valid.

2. Menghitung Tingkat Kesukaran

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan

⁴ Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. (Bandung: Alfabeta, 2009), h. 225

⁵ Suharsimi Arikunto, *Op.cit*, h.253

⁶ Widoyoko, E.S. *Teknik Penyusunan Instrumen Penelitian*. (Yogyakarta: Pustaka Pelajar, 2012), h. 149

dalam bentuk proporsi yang besarnya berkisar 0,00 - 1,00.⁷ Semakin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu. Suatu soal memiliki TK= 0,00 artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki TK= 1,00 artinya bahwa siswa menjawab benar. Perhitungan indeks tingkat kesukaran ini dilakukan untuk setiap nomor soal. Pada prinsipnya, skor rata-rata yang diperoleh peserta didik pada butir soal yang bersangkutan dinamakan tingkat kesukaran butir soal itu.

Fungsi tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Misalnya untuk keperluan ujian semester digunakan butir soal yang memiliki tingkat kesukaran sedang, untuk keperluan seleksi digunakan butir soal yang memiliki tingkat kesukaran tinggi/sukar, dan untuk keperluan diagnostik biasanya digunakan butir soal yang memiliki tingkat kesukaran rendah/mudah. Menganalisis tingkat kesukaran soal adalah untuk menentukan kualitas soal yang baik, mengetahui klasifikasi soal mudah, sedang dan sukar. Rumus yang digunakan untuk soal obyektif adalah sebagai berikut:

$$\begin{aligned} \text{Tingkat Kesukaran (TK)} \\ = \frac{\text{Jumlah siswa yang menjawab benar}}{\text{Jumlah siswa yang mengikuti tes}} \dots^8 \end{aligned}$$

Selanjutnya untuk mengetahui tingkat kesukaran soal bentuk uraian digunakan rumus berikut:

$$\begin{aligned} \text{Rata - rata} &= \frac{\text{Jumlah skor siswa pada suatu soal}}{\text{Jumlah siswa yang mengikuti tes}} \text{ dan} \\ \text{Tingkat Kesukaran} &= \frac{\text{Rata - rata}}{\text{Skor maksimum yang ditetapkan}} \dots^9 \end{aligned}$$

Kriteria tingkat kesukaran suatu item soal dibuat klasifikasi, yaitu;

- 1) Indeks kesukaran 0,00-0,30 tergolong sukar,
- 2) Indeks kesukaran 0,31-0,70 tergolong sedang, dan
- 3) Indeks kesukaran 0,71-1,00 tergolong mudah.

⁷Aiken, Lewis R. *Psychological Testing and Assessment*, (Eight Edition), Boston: Allyn and Bacon, . 1994. h. 66

⁸ Nitko, Anthony J. *Educational Assessment of Students*, (Second Edition. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs, 1996) h. 310

⁹ *Ibid*, h. 311

Butir soal dengan kategori mudah dan sukar dibuang.¹⁰ Berdasarkan hasil perhitungan indeks kesukaran, maka kemungkinan tidak semua soal dapat memenuhi kriteria pengujian. Soal yang mempunyai indeks kesukaran sedang/cukup yang dapat diambil.

3. Menghitung Daya Pembeda

Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan antara warga belajar/siswa yang telah menguasai materi yang ditanyakan dan warga belajar/siswa yang tidak/kurang/belum menguasai materi yang ditanyakan. Manfaat daya pembeda butir soal adalah; 1) untuk meningkatkan mutu setiap butir soal melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak, 2) untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru.

Apabila suatu butir soal tidak dapat membedakan kedua kemampuan siswa, maka butir soal itu dapat diduga bahwa kemungkinannya sebagai berikut:

- ✓ Kunci jawaban butir soal itu tidak tepat.
- ✓ Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar
- ✓ Kompetensi yang diukur tidak jelas
- ✓ Pengecoh tidak berfungsi
- ✓ Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak
- ✓ Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya

Untuk mengetahui daya pembeda tes pilihan ganda adalah dengan menggunakan rumus berikut ini.

$$DP = \frac{BA - BB}{\frac{1}{2}N} \text{ atau } DP = \frac{2(BA - BB)}{N} \dots^{11}$$

Keterangan:

DP = Daya pembeda soal,

BA = Jumlah jawaban benar pada kelompok atas,

BB = Jumlah jawaban benar pada kelompok bawah,

N = Jumlah siswa yang mengerjakan tes.

¹⁰ Suharsimi Arikunto, *Op.cit*, h.210

¹¹ Glass and Stanley, *Statistical Methods in Education and Psychology*. (New Jersey: Prentice Hall, Inc. 1970), h 169

Untuk mengetahui daya pembeda tes bentuk uraian adalah dengan menggunakan rumus berikut:

$$DP = \frac{\text{Mean kelompok atas} - \text{Mean kelompok bawah}}{\text{Skor maksimum soal}} \dots^{12}$$

Hasil perhitungan dengan menggunakan rumus di atas dapat menggambarkan tingkat kemampuan soal dalam membedakan antar peserta didik yang sudah memahami materi yang diujikan dengan peserta didik yang belum/tidak memahami materi yang diujikan.

Kriteria indeks daya pembeda soal dibuat klasifikasi sebagai berikut:

0,40 - 1,00 soal diterima baik

0,30 - 0,39 soal diterima tetapi perlu diperbaiki

0,20 - 0,29 soal diperbaiki

0,19 - 0,00 soal tidak dipakai/dibuang.¹³

Setelah diadakan perhitungan daya beda maka dari sejumlah soal yang disusun kemungkinan tidak semuanya dapat terpakai. Soal yang dapat terpakai adalah soal yang mempunyai daya beda cukup (0,21-0,40), baik (0,41-0,71), dan baik sekali (0,71-1,00).¹⁴

4. Analisis Distraktor (Pengecoh)

Analisis distraktor diperlukan hanya untuk pembuat soal. Selain menghitung indeks kesukaran dan daya beda dalam analisis butir juga perlu diketahui apakah distraktor atau pengecoh yang disediakan tepat atau tidak benar. Apakah semua pilihan yang disediakan dipilih semua karena dianggap betul, jawaban terkumpul pada pilihan tertentu atau pilihan yang sama sekali tidak ada pemilihnya. Dengan memeriksa pola pilihan jawaban, dapat ditentukan hal-hal, yaitu; 1) Berapa jumlah subjek yang menjawab betul, 2) Distraktor mana yang terlalu jelas atau menyolok sehingga sangat sedikit yang terkecoh untuk memilihnya, 3) Distraktor mana yang justru menyesatkan subjek yang termasuk kelompok tinggi yang seharusnya tidak terkecoh, dan 4) Distraktor mana yang dapat menarik bagi subjek kelompok rendah, tetapi tidak cukup menarik bagi subjek dari kelompok tinggi. Pengecoh dikatakan efektif bila minimal dijawab oleh 5% peserta dan lebih banyak dipilih oleh kelompok siswa yang belum paham materi, berarti $5\% \times (A+B)$.

¹² *Ibid*, h.170

¹³ Crocker, L. & Algina, J. *Introduction to Classical and Modern Test, Theory...* (New York: Holt, Rinehart and Winston, Inc. 1986) h.315

¹⁴ Suharsimi Arikunto, *Opcit*, h. 13

5. Analisis Reliabilitas Tes

Tujuan utama menghitung reliabilitas skor tes adalah untuk mengetahui tingkat ketepatan (*precision*) dan keajegan (*consistency*) skor tes. Suatu instrument atau tes dikatakan memiliki nilai reliabilitas yang tinggi apabila tes yang dibuat memiliki hasil yang konsisten dalam mengukur yang hendak diukur.¹⁵ Indeks reliabilitas berkisar antara 0-1. Semakin tinggi koefisien reliabilitas suatu tes (mendekati 1), makin tinggi pula keajegan/ketepatannya.¹⁶

Tes yang memiliki konsistensi reliabilitas tinggi adalah akurat, *reproducibel*, dan *generalized* terhadap kesempatan *testing* dan instrumen tes lainnya. Secara rinci faktor yang dapat mempengaruhi reliabilitas sebuah tes di antaranya:

- a. Semakin banyak jumlah butir soal, semakin ajek suatu tes.
- b. Semakin lama waktu tes, semakin ajek.
- c. Semakin sempit range kesukaran butir soal, semakin besar keajegan.
- d. Soal-soal yang saling berhubungan akan mengurangi keajegan.
- e. Semakin objektif pemberian skor, semakin besar keajegan.
- f. Ketidaktepatan pemberian skor.
- g. Menjawab besar soal dengan cara menebak.
- h. Semakin homogen materi semakin besar keajegan.
- i. Pengalaman peserta ujian.
- j. Salah penafsiran terhadap butir soal.
- k. Menjawab soal dengan buru-buru/cepat.
- l. Kesiapan mental peserta ujian.
- m. Adanya gangguan dalam pelaksanaan tes.
- n. Jarak antara tes pertama dengan tes kedua.
- o. Mencontek dalam mengerjakan tes.
- p. Posisi individu dalam belajar.
- q. Kondisi fisik peserta tes/ujian.

Selanjutnya ada 3 (tiga) cara yang dapat dilakukan untuk menentukan reliabilitas skor tes, yaitu :

- a. Keajegan pengukuran ulang: kesesuaian antara hasil pengukuran pertama dan kedua dari sesuatu alat ukur terhadap kelompok yang sama.
- b. Keajegan pengukuran setara: kesesuaian hasil pengukuran dan 2 atau lebih alat ukur berdasarkan kompetensi kisi-kisi yang lama.

¹⁵ Sukardi. *Metodologi Penelitian Pendidikan: Kompetensi dan Praktiknya*. (Jakarta: Bumi Aksara, 2012), h. 127.

¹⁶ Depdiknas, *Juknis Penulisan Analisis Butir Soal di SMA*. (Jakarta: Direktorat Pembinaan SMA, 2010), h. 121.

c. Keajegan belah dua: kesesuaian antara hasil pengukuran belahan pertama dan belahan kedua dari alat ukur yang sama.

Untuk mengetahui koefisien reliabilitas tes/soal bentuk pilihan ganda digunakan rumus Kuder Richadson 21 (KR-21) dengan persamaan:

$$r_{11} = \frac{n}{n-1} \left[1 - \frac{\sum \bar{x}(n-\bar{x})}{n(SD)^2} \right] \dots^{17}$$

Keterangan:

r_{11} = koefisien reliabilitas
 n = Banyaknya butir soal
 \bar{x} = Rata-rata skor total
 $(SD)^2$ = Varians skor total

Untuk menghitung reliabilitas tes bentuk uraian dapat dilakukan dengan menggunakan rumus Cronbach-Alpha, yaitu:

$$r_{11} = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n s_i^2}{s_t^2} \right)$$

Keterangan :

r_{11} = koefisien reliabilitas
 n = banyaknya butir soal
 s_i^2 = varians skor soal ke-i
 s_t^2 = varians skor total

Nilai korelasi (r) yang diperoleh kemudian dikonsultasikan dengan nilai tabel koefisien korelasi dengan kategori sebagai berikut:

- $0,90 \leq r \leq 1,00$ = korelasi sangat tinggi (sempurna)
- $0,70 \leq r \leq 0,90$ = korelasi tinggi
- $0,40 \leq r \leq 0,70$ = korelasi sedang
- $0,20 \leq r \leq 0,40$ = korelasi rendah
- $0,00 \leq r \leq 0,20$ = korelasi sangat rendah.¹⁹

Jika hasil perhitungan diperoleh besarnya $r=0,76$, maka berarti $0,76$ terletak antara $0,6 \leq r \leq 0,8$, berarti koefisien reliabilitas tes tergolong kategori tinggi.

¹⁷ Glass and Stanley. *Statistical Methods in Education and Psychology*.(New Jersey: Prentice Hall, Inc. 1970) . h 172

¹⁸ *Ibid*, h.173

¹⁹ Depdiknas, *Opcit*, h. 129

Berdasarkan uji metode belah duaini, maka tes yang dicari reliabilitasnya adalah tes yang sah saja, yaitu tes yang sudah valid atau memenuhi ke tiga 3 kreteria analisis butir (r butir, indeks kesukaran, daya pembeda).Berdasarkan hal tersebut maka tidak semua soal dapat dipakai, untuk itu perlu kembali penatan ulang nomor soal.

D. PENUTUP

1. Dalam penyusunan tes hasil belajar perlu menyiapkan bahan atau materi soal terkait diantaranya kurikulum, silabus dan SAP dengan memperhatikan kompetensi inti, kompetensi dasar, pengecekan kesesuaian indikator, pembuatan indikator soal. Hal ini diharapkan agar tidak terjadi kesalahan atau ketidaktepatan dalam pembuatan butir soal.
2. Pembuatan kisi-kisi sangat penting artinya dalam penyusunan soal agar mendapatkan soal-soal yang mewakili isi kurikulum
3. Sukar atau mudahnya suatu soal bukan semata-mata ditentukan oleh materi soal, akan tetapi ditentukan juga oleh teknik penyusunannya.
4. Soal yang memenuhi persyaratan adalah soal yang valid dan reliabel. Maka perlu dilakukan analisis butir soal dengan menganalisis korelasi antar butir soal, tingkat kesukaran, daya pembeda, fungsi distractor (pengecoh).

DAFTAR PUSTAKA

- Aiken, Lewis R. *Psychological Testing and Assessment*,(Eight Edition), Boston: Allyn and Bacon, 1994
- Arikunto,Suharsimi. *Dasar-Dasar Evaluasi Pendidikan*, Jakarta: Bumi Aksara, 2003
- Azwar, Syarifuddin. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*, edisi II, cetakan ke 4, 2003
- Crocker, L. & Algina, J. *Introduction to Classical and Modern Test, Theory_*. New York: Holt, Rinehart and Winston, Inc, 1986
- Depdiknas, *Juknis Penulisan Analisis Butir Soal di SMA*. Jakarta: Direktorat Pembinaan SMA, 2010
- Glass and Stanley. *Statistical Methods in Education and Psychology*.New Jersey: Prentice Hall, Inc, 1970
- Nitko, Anthony J. *Educational Assessment of Students*, Second Edition. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs, 1996
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*, Jakarta: Rajawali Press, 2000
- Sudjana, Nana dan Rivai, A. *Teknologi Pembelajaran*. Bandung: Sinar Baru Algesindo 2001

- Sugiyono. *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2009
- Sukardi. *Metodologi Penelitian Pendidikan: Kompetensi dan Praktiknya*. Jakarta: Bumi Aksara, 2012
- Thoha, M. Chabib. *Teknik Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada, 2001
- Widoyoko, E.S. *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar, 2012.
- Winkel, W.S. *Psikologi Pembelajaran*. Jakarta : PT. Gramedia, 1989