

Tahapan Penyusunan dan Analisis Tes Bahasa Arab

Melisa Rezi¹, Annisa Aulia²

Abstrak

Tes Bahasa Arab merupakan salah satu dari alat evaluasi pembelajaran Bahasa Arab, di mana evaluasi merupakan salah satu komponen penting dalam kurikulum pembelajaran Bahasa Arab. Untuk mendapatkan hasil tes yang baik, bermakna, dan bermutu, maka perlu adanya perencanaan tes Bahasa Arab sesuai dengan tahapan penyusunannya, yaitu di antaranya: (Persiapan, Pemilihan materi tes, Menentukan bentuk dan jenis tes, Menentukan jumlah butir tes, Menentukan skor, Membuat kisi-kisi, Menyusun butir tes berdasarkan kisi-kisi, Uji coba tes yang telah disusun), dan juga perlu adanya analisa terhadap tes Bahasa Arab yang sudah direncanakan dan diujicobakan terhadap siswa baik itu menganalisa Tingkat kesukaran soal, Daya Pembeda dan fungsi distractor, sehingga hasil tes bahasa Arab siswa yang kita dapatkan itu valid dan reliabel.

Kata Kunci: Tahapan Penyusunan Tes, Analisis Tes, Tes Bahasa Arab

Abstract

Arabic test is one of the tools for Arabic learning evaluation, where the evaluation is one of the critical components in the Arabic learning curriculum. To get good, meaningful, and meaningful test results, Arab test planning must be consistent with the actual point of composition: (preparation, selection of test materials, determining the number of test items, determining scores, creating lattices, compiling of test items based on lattices, testing of tests compiled), and also requiring an analysis of the conceived and assigned Arabic tests analyzing the level of difficulty, differential and function of the tors, so that the student's Arabic test is valid and religious.

Keywords: photodrafting sequence, test analysis, Arabic test

PENDAHULUAN

Melihat fenomena yang kebanyakan terjadi, bahwasannya murid atau pelajar yang mempelajari bahasa asing dalam hal ini bahasa arab sangat sulit untuk bisa memahami dengan betul makna dan tujuan bahasa tersebut, sehingga mereka semua tidak bisa menerapkan hasil pembelajaran tersebut dalam kehidupan sehari-hari. Memang tidak bisa dipungkiri belajar bahasa arab merupakan suatu kesulitan tersendiri, tidak lain karena kita dihadapkan dengan adanya bahasa ibu yang sudah dipergunakan sejak lahir, setiap hari dan setiap saat. Untuk bisa memahami bahasa arab perlu tahapan-tahapan yang prosesnya sangat lama dan dalam tahapan tersebut membutuhkan proses pendidikan yang *continue* sekaligus terarah. Tahapan tersebut antara lain adalah tahap pengenalan, pendengaran, pengucapan, dan

¹ Dosen Tetap STAIDA Payakumbuh, melisarezi10@gmail.com

²Guru Bahasa Arab Pondok Pesantren MTI Canduang, pramicha21@gmail.com

pembiasaan. Masalah yang lain adalah bahasa arab itu berbeda dengan bahasa ibu yakni dalam hal segi-segi suara, kosakata, tata kalimat dan tulisan, sehingga si pembelajar harus berjuang untuk memahami perbedaan tersebut dengan detail.

Dengan adanya tahapan-tahapan dan perbedaan di atas, seorang guru bahasa arab harus berperan penuh demi pencapaian kompetensi ketrampilan berbahasa. Ini menjadi tantangan besar bagi para pahlawan tanpa jasa tersebut. Melihat permasalahan yang sering terjadi pada guru bahasa arab yakni mereka tidak menghiraukan apa yang dinamakan “perencanaan pembelajaran”, mereka asal-asalan dalam mengajar dengan berdalih yang penting materinya selesai. Proses pembelajaran yang terakhir disebut dengan evaluasi belajar. Di sini guru harus bisa membuat alat untuk mengevaluasi hasil belajar yang sudah direncanakan. Salah satu alat evaluasi tersebut adalah tes. Seorang guru bahasa harus benar-benar mengetahui perencanaan pembuatan tes bahasa untuk muridnya demi tercapainya indikator keberhasilan yang sudah ditentukan. Oleh karena itulah, demi pemahaman yang sempurna, suatu tes bahasa harus direncanakan terlebih dahulu sebelum dilaksanakan. Karena dengan adanya perencanaan pembuatan tes bahasa, guru mampu mengevaluasi atau melaksanakan tes bahasa dengan cara atau prosedur yang baik dan benar. Dan dengan tes yang baik dan benar tersebut, dimungkinkan indikator-indikator pencapaian belajar bisa berhasil.

Tes bahasa arab yang baik adalah tes yang memenuhi standar validitas, reliabilitas, dan memiliki tingkat kesulitan dan daya beda yang baik. Untuk menghasilkan tes bahasa arab yang memenuhi standar tersebut, ada tahapan-tahapan prosedural yang harus diperhatikan oleh penyusun tes. Tahapan tersebut meliputi: tahapan persiapan, pemilihan materi tes, penentuan bentuk tes, penentuan jumlah butir soal, pembuatan kisi-kisi, penyusunan tes(butir soal), uji coba, dan analisis hasil uji coba yang mencakup analisis tingkat kesulitan, daya beda, dan reliabilitas.

PEMBAHASAN

A. Prinsip Prinsip Dalam Penyusunan Tes Hasil Belajar

Ada beberapa prinsip dasar yang perlu dicermati dalam menyusun tes hasil belajar agar tes tersebut dapat mengukur tujuan dari mata pelajaran yang telah diajarkan, atau mengukur kemampuan dan keterampilan peserta didik yang diharapkan, setelah mereka menyelesaikan satu unit pengajaran tertentu. Adapun prinsip tersebut adalah :

1. Tes hasil belajar harus dapat mengukur secara jelas hasil belajar yang telah ditetapkan sesuai dengan tujuan. Kejelasan mengenai pengukuran hasil belajar yang dikehendaki akan memudahkan bagi guru dalam menyusun butir-butir soal tes hasil belajar.
2. Butir-butir soal tes hasil belajar harus merupakan sampel yang representatif dari populasi bahan pelajaran yang telah diajarkan, sehingga dapat dianggap mewakili seluruh performance yang telah diperoleh selama peserta didik mengikuti suatu unit pengajaran.
3. Bentuk soal yang dikeluarkan dalam tes hasil belajar harus dibuat bervariasi, sehingga betul-betul cocok untuk mengukur hasil belajar yang diinginkan sesuai dengan tujuan tes itu sendiri. Untuk mengukur hasil belajar yang berupa keterampilan misalnya, tidak tepat kalau hanya menggunakan soal-soal yang berbentuk essay tes yang jawabannya hanya menguraikan dan bukan melakukan atau mempraktekkan sesuatu. Demikian pula untuk mengukur kemampuan menganalisis suatu prinsip, tidak cocok jika digunakan butir-butir soal yang berbentuk objektif tes yang pada dasarnya hanya mengungkap daya ingat peserta didik.
4. Tes hasil belajar harus di desain sesuai dengan kegunaannya untuk memperoleh hasil yang diinginkan. Pernyataan tersebut mengandung makna, bahwa desain tes hasil belajar harus disusun relevan dengan kegunaan dimiliki oleh masing-masing jenis tes. Desain dari placement tes (yaitu tes yang digunakan untuk penentuan penempatan siswa dalam suatu jenjang atau jenis program pendidikan tertentu), sudah barang tentu akan berbeda dengan desain dari formatif tes (tes yang digunakan untuk mencari umpan balik guna memperbaiki proses pembelajaran, baik bagi guru maupun bagi siswa), dan tes sumatif (tes yang digunakan untuk mengukur atau menilai sampai dimana pencapaian siswa terhadap bahan pelajaran yang telah diajarkan dan selanjutnya untuk menentukan kenaikan tingkat atau kelulusan siswa yang bersangkutan). Demikian pula desain diagnostik tes (yaitu tes yang digunakan dengan tujuan untuk mencari sebab-sebab kesulitan belajar siswa, seperti latar belakang psikologis, fisik, dan lingkungan sosial ekonomi siswa), tentu akan berbeda pula dengan dengan tiga jenis tes yang telah disebutkan diatas.
5. Tes hasil belajar harus memiliki reliabilitas yang dapat diandalkan. Artinya, setelah tes hasil belajar itu dilaksanakan berkali-kali terhadap subjek yang sama, hasilnya

selalu sama atau relatif sama. Dengan demikian tes hasil belajar itu hendaknya memiliki keajegan hasil pengukuran yang tidak diragukan lagi.

6. Tes hasil belajar disamping harus dapat dijadikan alat pengukur keberhasilan belajar siswa, juga harus dapat dijadikan alat untuk mencari informasi yang berguna untuk memperbaiki cara belajar siswa dan cara mengajar guru itu sendiri.³

B. Tahapan Penyusunan Soal (Tes) Bahasa Arab

Sebelum membicarakan masalah teknik penyusunan dan pelaksanaan tes hasil belajar, alangkah baiknya mengenali dulu ciri-ciri tes hasil belajar yang baik, sehingga diharapkan nantinya dalam menyusun dan melaksanakan tes hasil belajar tidak mengalami kesulitan dalam menentukan soal-soal yang akan diteskan.

Setidak-tidaknya ada empat ciri atau karakteristik yang harus dimiliki oleh tes hasil belajar, sehingga tester tersebut dapat dinyatakan sebagai tes yang baik, yaitu : (1) valid (shahih/tepat), (2) reliabel (andal/tetap/ajeg), (3) obyektif dan (4) praktis.

Ciri pertama, dari tes hasil belajar yang baik adalah bahwa tes hasil belajar tersebut bersifat valid atau memiliki validitas. Kata valid sering diartikan dengan tepat, benar, shahih, absah. Jadi kata validitas dapat diartikan dengan ketepatan, kebenaran, kesahihan atau keabsahan. Apabila kata valid itu dikaitkan dengan fungsi tes sebagai alat pengukur, maka sebuah tes dikatakan telah memiliki "validitas" apabila tes tersebut dapat tepat mengukur apa yang hendak diukur. Contoh, untuk mengukur partisipasi siswa dalam proses belajar mengajar, bukan diukur melalui nilai yang diperoleh pada waktu ulangan, tetapi dilihat melalui: kehadiran, terpusatnya perhatian pada pelajaran, ketepatan menjawab pertanyaan-pertanyaan yang diajukan oleh guru dalam arti relevan pada permasalahannya.⁴ Jadi tes hasil belajar dapat dinyatakan valid apabila hasil belajar tersebut (sebagai alat pengukur keberhasilan belajar peserta didik) dengan secara tepat benar, shahih, atau absah dapat mengukur atau mengungkap hasil-hasil belajar yang telah dicapai oleh peserta didik, setelah menempuh proses pembelajaran dalam jangka waktu tertentu.

³Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta : PT Raja Grafindo Persada, 2003), hal. 97-99

⁴Fatimah Depi Susanty, 2016, *Analisis Validasi Soal Tes Hasil Belajar Pada Pelaksanaan Pembelajaran Bahasa Arab Di Pusat Pengembangan Bahasa (P3b) Uin Suska Riau*, Kutubkhanah: Jurnal Penelitian sosial keagamaan, Vol.19, No.2 Juli -Desember

Ciri kedua, dari tes hasil belajar yang baik adalah bahwa tes hasil belajar tersebut telah memiliki reliabilitas atau bersifat reliabel. Kata "reliabilitas" sering diterjemahkan dengan keajegan (stability) atau kemantapan (consistency). Apabila istilah tersebut dikaitkan dengan fungsi peserta didik, maka sebuah tes hasil belajar dapat dinyatakan reliabel apabila hasil-hasil pengukuran yang dilakukan dengan menggunakan tes tersebut secara berulang kali terhadap subyek yang sama, senantiasa menunjukkan hasil yang tetap sama atau sifatnya tetap dan stabil. Dengan demikian suatu ujian dikatakan telah memiliki reliabilitas (daya keajegan mengukur) apabila skor-skor atau nilai-nilai yang diperoleh para peserta ujian adalah stabil, kapan saja, dimana saja dan oleh siapa saja ujian itu dilaksanakan, diperiksa dan dinilai.

Dari uraian di atas dapat dipahami bahwa prinsip reliabilitas akan menyangkut pertanyaan: "seberapa jauhkah pengukuran yang dilakukan secara berulang kali terhadap subyek atau kelompok subyek yang sama, memberikan hasil-hasil yang relatif tidak mengalami perubahan."

Ciri ketiga dari tes hasil belajar yang baik adalah, bahwa tes hasil belajar tersebut bersifat obyektif. Dalam hubungan ini sebuah tes hasil belajar dapat dikatakan sebagai tes hasil belajar yang obyektif, apabila tes tersebut disusun dan dilaksanakan "menurut apa adanya". Ditinjau dari segi isi atau materi tesnya, maka istilah "apa adanya, itu mengandung pengertian bahwa materi tes tersebut adalah diambilkan atau bersumber dari materi atau bahan pelajaran yang telah diberikan, sesuai atau sejalan dengan tujuan pembelajaran atau indikator atau hasil belajar yang telah ditentukan di dalam kurikulum. Dilihat dari segi pemberian skor dan penentuan nilai hasil tesnya, maka dengan istilah "apa adanya" itu terkandung pengertian bahwa pekerjaan koreksi, pemberian skor dan penentuan nilainya terhindar dari unsur-unsur subyektivitas yang melekat pada diri penyusun tes. Di sini tester harus bisa meminimalisir sejauh mungkin kemungkinan kemunculan "hallo effect" seperti jawaban soal dengan tulisan yang baik mendapat skor lebih tinggi daripada jawaban soal yang tulisannya jelek, padahal jawaban tersebut sama. Demikian pula kesan rasa malu atau kasihan dan lain-lain harus disingkirkan jauh-jauh sehingga tes hasil belajar tersebut menghasilkan nilai-nilai yang obyektif.

Ciri keempat dari tes hasil belajar yang baik adalah, bahwa tes hasil belajar tersebut bersifat praktis (practicability) dan ekonomis. Bersifat praktis mengandung

pengertian bahwa tes hasil belajar tersebut dapat dilaksanakan dengan mudah karena itu : (a) bersifat sederhana, dalam arti tidak memerlukan peralatan yang banyak atau peralatan yang sulit pengadaannya, (b) lengkap, dalam arti bahwa tes tersebut telah dilengkapi dengan petunjuk mengenai cara mengerjakannya kunci jawabannya dan pedoman skoring serta penentuan nilainya. Bersifat ekonomis mengandung pengertian bahwa tes hasil belajar tersebut tidak memakan tenaga biaya dan waktu yang banyak.⁵

Dengan memperhatikan karakteristik tersebut, maka tes yang disusun benar-benar mampu mengukur yang seharusnya diukur. Permasalahannya adalah seringkali tes yang disusun itu kurang memperhatikan unsur validitas dan reliabilitasnya. Implikasinya, sering dijumpai bahwa tes yang disusun terlalu sulit atau terlalu mudah, atau bahkan tes yang disusun tidak mencerminkan kemampuan yang seharusnya diukur.⁶

Untuk menghasilkan suatu tes yang valid dan reliable, maka pembuat tes (guru) harus menempuh langkah-langkah sebagai berikut :

1. Tahap Persiapan

Pada tahap ini guru atau pembuat tes melakukan kajian terhadap kurikulum bahasa Arab dan buku pedoman pelaksanaan kurikulum untuk mata pelajaran bahasa Arab. Apabila kurikulum yang dijadikan sandaran adalah kurikulum 2004 Berbasis Kompetensi, maka substansi yang dikaji meliputi Kompetensi Dasar, Indikator, hasil, topik-topik bahasan, penilaian, dan alokasi waktu yang tersedia.⁷

2. Pemilihan Materi Tes

Untuk menetapkan materi tes bahasa Arab yang benar-benar *fixed* dan selektif dapat dilakukan beberapa kegiatan sebagai berikut :

- a. Menentukan komponen dan keterampilan berbahasa yang akan diteskan, misalnya tes kosakata, struktur, membaca, menulis atau tes berbicara.
- b. Menentukan pokok bahasan yang akan diteskan secara representatif (tidak bias, dan tidak atas dasar subjektif penyusunan tes).⁸

3. Menentukan Bentuk Dan Jenis Tes

⁵Gito Supriadi, *Pengantar dan Teknik Evaluasi Pembelajaran*, (Malang : Intimedia Press, 2011), hal. 35-38

⁶M. Ainin, Dkk, *Evaluasi Dalam Pembelajaran Bahasa Arab*, (Malang : MISYKAT, 2006), hal. 94

⁷Yelfi Dewi, *Evaluasi dan tes Bahasa Arab Pedoman bagi Guru, Mahasiswa dan Praktisi Pendidikan Bahasa Arab*, (Bukittinggi : STAIN Bukittinggi Press, 2013), hal. 45

⁸M. Ainin, Dkk, *Evaluasi Dalam Pembelajaran Bahasa Arab*, hal. 95

Sebagaimana telah dikemukakan, tes komponen bahasa dan kemampuan bahasa dapat disusun dalam bentuk subjektif atau objektif dengan segala variasinya atau jenisnya (kecuali tes keterampilan berbicara yang memiliki perlakuan khusus. Dengan ungkapan lain, tes yang disusun dapat berbentuk objektif dengan jenis pilihan ganda atau salah benar, atau dapat pula berbentuk subjektif (*essay*).

4. Menentukan Jumlah Butir Tes

Perihal yang harus dipertimbangkan dalam menentukan jumlah butir tes adalah alokasi waktu yang tersedia untuk penyelenggaraan tes. Untuk menentukan berapa jumlah butir tes yang harus disusun sesuai dengan waktu yang tersedia memang tidak ada batasan yang pasti. Akan tetapi, guru dengan nalurinya yang mengetahui kondisi objektif siswanya akan dapat menentukan jumlah butir tes yang sesuai dengan waktu yang tersedia. Misalnya satu nomor soal dalam pilihan ganda diberikan batas waktu satu menit.

Berkaitan dengan hubungan penentuan jumlah butir tes dan alokasi waktu yang tersedia, seorang guru atau pembuat tes perlu juga memperhatikan bentuk tesnya itu sendiri. Sudah barang tentu waktu yang digunakan untuk menjawab soal dalam bentuk esai lebih banyak daripada untuk menjawab soal dalam bentuk pilihan ganda atau salah benar.

Misalnya, jika jumlah butir tes itu 25, maka proporsi jumlah item untuk masing-masing jenis tes adalah 10 item untuk tes pilihan ganda, 10 item untuk tes salah benar, dan 5 item untuk tes esai. Demikian pula, proporsi jumlah butir tes untuk masing-masing sub kemampuan juga perlu diperhatikan.

Dalam kaitannya dengan penentuan jumlah butir tes untuk masing-masing aspek (kosakata, struktur, membaca, dan menulis), pembuat tes dapat merumuskannya dengan pertimbangan tertentu. Diantaranya dengan menghitung alokasi waktu yang tersedia untuk mata pelajaran bahasa Arab secara keseluruhan dan alokasi waktu untuk masing-masing aspek. Jika dalam kurikulum tidak disebutkan alokasi waktu untuk masing-masing komponen atau keterampilan, maka jumlah butir tes dapat ditentukan melalui butir-butir tujuan pembelajaran atau prioritas tujuan pembelajaran.⁹

Dalam contoh yang lain, seandainya alokasi waktu yang tersedia 90 menit, kemudian pembuat tes menetapkan jumlah butir tes sebanyak 40 dan disusun

⁹Moh. Matsna HS, *Pengembangan Evaluasi dan Tes Bahasa Arab*, (Tangerang : Alkitabiah, 2012), hal. 77-78

dalam bentuk pilihan ganda maka rincian perkiraan penggunaan waktu 90 menit tersebut adalah sebagai berikut :

- a. 5 menit digunakan oleh guru atau petugas untuk memberikan arahan atau penjelasan sekilas tentang penyelenggaraan tes.
- b. 1 menit digunakan oleh teste (siswa) untuk membaca petunjuk mengerjakan soal.
- c. 1,5 menit digunakan oleh siswa untuk menjawab setiap butir tes kosakata yang jumlah butir tesnya sebanyak 9 ($9 \times 1,5 \text{ menit} = 13,5 \text{ menit}$).
- d. 1,5 menit digunakan oleh siswa untuk menjawab setiap butir tes tata bahasa yang jumlah butir tesnya sebanyak 9 ($9 \times 1,5 \text{ menit} = 13,5 \text{ menit}$).
- e. 2,57 menit digunakan oleh siswa untuk menjawab setiap butir tes membaca yang jumlah butir tesnya sebanyak 15 ($15 \times 2,57 \text{ menit} = 38,5 \text{ menit}$).
- f. 2,64 menit digunakan oleh siswa untuk menjawab setiap butir tes menulis terbimbing yang jumlah butir tesnya sebanyak 7 ($7 \times 2,64 \text{ menit} = 18,5 \text{ menit}$).

Adapun cara menentukan jumlah butir tes adalah sebagai berikut, sebagai contoh, alokasi waktu mata pelajaran bahasa Arab untuk kelas XII Madrasah Aliyah (semester I dan II) misalnya 48,6 jam sudah dikurangi alokasi waktu untuk berbicara), sedangkan alokasi waktu untuk kosakata adalah 10,8 jam, struktur 10,8 jam, membaca 18 jam, dan menulis 8 jam, maka dapat digunakan rumus sebagai berikut :

Rumus I :

$$\frac{\sum \text{jam per aspek}}{\sum \text{jam keseluruhan}} \times 100$$

Dengan demikian :

- b. Kosa kata $: \frac{10,8}{48,6} \times 100 = 22,22$
- c. Struktur $: \frac{10,8}{48,6} \times 100 = 22,22$
- d. Membaca $: \frac{18}{48,6} \times 100 = 37,03$
- e. Menulis $: \frac{8}{48,6} \times 100 = 18,52$

Rumus II :

$$\frac{\text{Hasil Perhitungan Rumus I}}{100} \times \text{jumlah soal}$$

Dengan demikian :

- a. Kosakata : $\frac{22}{100} \times 40 = 8,89$ (9 butir tes)
- b. Struktur : $\frac{22}{100} \times 40 = 8,89$ (9 butir tes)
- c. Membaca : $\frac{3707}{100} \times 40 = 14,81$ (15 butir tes)
- d. Menulis : $\frac{18,52}{100} \times 40 = 7,41$ (7 butir tes)¹⁰Type equation here.

5. Menentukan Skor

Apabila jumlah butir tes sebanyak 40 (pilihan ganda) dengan skor tertinggi 100, dan semua butir tes diberi bobot skor sama, maka skor untuk jawaban yang benar pada setiap butir tes adalah 2,5.

Sementara itu, jawaban tes esay dapat di skor sesuai dengan pendapat dan penilaian subjektif (berbasis keahlian profesional) seorang korektor. Jika seandainya suatu pekerjaan tes subjektif diperiksa oleh dua orang atau lebih korektor yang berbeda, maka hasil penilaiannya sangat mungkin akan berbeda antara satu korektor dengan korektor lainnya. Perbedaan itu disebabkan oleh pendapat penilaian masing-masing korektor yang subjektif terhadap pekerjaan yang sama dari peserta tes yang sama.

Apabila soal yang dibuat mempunyai tingkat kesukaran yang berbeda-beda, maka pihak pembuat tes dapat memberikan bobot yang berbeda. Artinya, suatu butir tes diberi bobot tinggi apabila butir tes tersebut lebih sulit dan lebih kompleks bila dibandingkan dengan butir tes lain.

Dengan mengingat kedua jenis tes tersebut kita dapat menjelaskan perbedaan pada pemberian skornya, yaitu sebagai berikut :

- a. Untuk tes objektif dalam menentukan benar dan salahnya jawaban terhadap suatu butir tes penilai hanya mencocokkan jawaban dengan kunci jawaban yang tersedia. Sedangkan untuk tes subjektif penilai seringkali masih perlu mempertimbangkan betul salahnya suatu jawaban yang tidak persis sama dengan yang tertulis pada kuncinya.
- b. Sebagai akibat dari perlunya pertimbangan penilai terhadap suatu jawaban, maka otomatis proses pemberian skor untuk tes subjektif akan lebih lama dibandingkan dengan proses pemberian skor pada tes objektif.

¹⁰M. Ainin, Dkk, *Evaluasi Dalam Pembelajaran Bahasa Arab*, hal. 96-98

- c. Jawaban terhadap tes objektif bersifat pasti, harus persis dengan kuncinya, maka korektornya tidak harus seseorang yang ahli di bidang yang diteskan. Sedangkan untuk tes subjektif, diperlukan korektor yang menguasai bidang yang diteskan.
- d. Siapapun yang mengoreksi tes objektif hasil skornya tidak akan bervariasi, sedangkan pada tes subjektif berbeda korektor mungkin saja menghasilkan skor yang berbeda.

Dari uraian diatas dapat disimpulkan bahwa pemberian skor untuk tes subjektif lebih kompleks dan lebih rumit, sedangkan untuk tes objektif lebih mudah dan lebih sederhana.¹¹

6. Membuat Kisi-kisi

Tujuan pembuatan kisi-adalah membuat spesifikasi yang jelas tentang apa yang akan ditanyakan. Tujuan membuat kisi-kisi adalah supaya apa yang akan kita usahakan mengenai sasaran. Kisi-kisi yang baik harus mewakili isi kurikulum suatu lembaga yang bersangkutan, komponen-komponen nya diuraikan dengan jelas dan mudah dipahami serta memberikan indikator yang jelas pula.¹²

Selain itu juga harus memperhatikan hal-hal berikut :

- a. Waktu untuk melaksanakan tes
- b. Topik-topik yang akan ditanyakan
- c. Jenis soalnya yang akan digunakan
- d. Jumlah soal setiap soal secara keseluruhan maupun tiap-tiap topik
- e. Persentase soal yang mudah dan yang sulit setelah diuji cobakan

Dalam membuat kisi-kisi soal, dapat dilakukan langkah-langkah berikut :

- a. Menuliskan kompetensi dasar dan indikator capaian hasil belajar yang terdapat dalam silabus/kurikulum
- b. Menuliskan daftar pokok/sub pokok bahasan yang akan diujikan
- c. Menentukan jumlah butir soal setiap pokok/sub pokok bahasan, jumlah soal hendaknya representatif untuk untuk setiap pokok /sub pokok bahasan yang diujikan dengan pertimbangan pentingnya pokok/sub pokok bahasan tersebut. Selain itu, dalam menentukan jumlah soal juga perlu mempertimbangkan waktu yang tersedia untuk pelaksanaan tes.

¹¹Ertah Mahyudin, *Pengembangan Evaluasi dan Tes Bahasa Arab*, (Tangerang : Alkitabiah, 2012), hal. 78 & 210

¹²Moh. Matsna HS, *Pengembangan Evaluasi dan Tes Bahasa Arab*..... hal. 79

- d. Menentukan bentuk soal tes, dalam menentukan bentuk tes, perlu mempertimbangkan karakteristik materi yang hendak diukur. Jika tes itu untuk mengukur pemahaman mufradat, qawaid, istima', dan qiraah, maka bisa digunakan bentuk tes pilihan ganda. Namun jika tes itu untuk mengukur kemampuan menulis dan berbicara, maka lebih tepat jika menggunakan bentuk tes uraian atau mengarang.¹³

7. Menyusun Butir Tes Berdasarkan Kisi-Kisi

Dalam penyusunan butir soal ini, ada rambu-rambu yang sebaiknya diperhatikan oleh guru atau pembuat tes, yaitu :

- a. Bahasa yang digunakan jelas dan lugas
- b. Stem (pernyataan pokok) pada setiap butir tes (pilihan ganda) hanya berisi satu permasalahan
- c. Panjang jawaban untuk setiap option (khusus untuk pilihan ganda) relatif sama
- d. Letak jawaban yang benar disusun secara acak, artinya harus dihindari letak jawaban benar yang berpola, misalnya berpola ab, ac, dan ad atau berpola aa, bb, cc, dan dd.

8. Uji Coba Tes yang telah disusun

Idealnya sebelum tes diberlakukan kepada siswa, perlu dilakukan uji coba terlebih dahulu. Uji coba ini bertujuan untuk mengetahui apakah tes yang disusun benar-benar tes yang baik atau apakah tes yang disusun itu memiliki tingkat kesulitan yang normal dan benar-benar dapat membedakan kelompok teste yang memiliki kemampuan tinggi dan rendah. Untuk mengetahui hal itu, maka setelah tes uji coba dilakukanlah analisis terhadap jawaban siswa. Diantara variabel yang dianalisis adalah analisis tingkat kesulitan, analisis daya beda, dan analisis reliabilitas.¹⁴

C. Analisis Hasil Tes Bahasa Arab

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. Tes sebagai alat evaluasi diharapkan menghasilkan nilai yang objektif dan akurat. Jika tes yang digunakan guru kurang baik, maka hasil yang

¹³Yelfi Dewi, *Evaluasi dan tes Bahasa Arab Pedoman bagi Guru, Mahasiswa dan Praktisi Pendidikan Bahasa Arab*.....hal. 47-48

¹⁴M. Ainin, Dkk, *Evaluasi Dalam Pembelajaran Bahasa Arab*, hal. 100-102

diperolehpun tentunya kurang baik. Hal ini dapat merugikan peserta didik itu sendiri. Artinya, hasil yang diperoleh peserta didik menjadi tidak objektif dan tidak adil. Oleh sebab itu, tes yang digunakan guru harus memiliki kualitas yang lebih baik dilihat dari berbagai segi. Tes hendaknya disusun sesuai dengan prinsip dan prosedur penyusunan tes. Setelah digunakan perlu diketahui apakah tes tersebut berkualitas baik atau kurang baik. Untuk mengetahui apakah suatu tes yang digunakan termasuk baik atau kurang baik, maka perlu dilakukan analisis kualitas tes.¹⁵

Penganalisisan terhadap butir-butir item tes hasil belajar meliputi :

1. Tingkat Kesukaran/kesulitan

Analisis tingkat kesulitan butir tes dimaksudkan untuk mengetahui seberapa sulit atau mudahnya tes yang telah diselenggarakan, baik tes secara keseluruhan maupun masing-masing butir tesnya. Tingkat kesulitan itu diperhitungkan dari perbandingan antara jumlah peserta tes yang dapat menjawab dengan benar dan yang tidak mampu menjawab dengan benar. Dasarnya adalah bahwa semakin banyak peserta tes yang menjawab dengan benar, semakin mudah tes atau butir tes yang bersangkutan. Oleh sebab itu, rumus sederhana untuk menghitung tingkat kesulitan butir tes berupa :

$$P = \frac{JJB}{JPT} \times 100\%$$

Keterangan :

P : Tingkat Kesulitan Butir Tes

JJB : Jumlah Jawaban Benar

JPT : Jumlah Peserta Tes¹⁶

Mengenai bagaimana cara memberikan penafsiran terhadap angka indekskesukaran item , Robert L. Thorndike dan Elizabeth Hagen dalam bukunya berjudul *Measurement and Evaluation in Psychology and Education* mengemukakan sebagai berikut :

Besarnya P	Interpretasi
Kurang dari 0,30	Terlalu Sukar
0,30 – 0,70	Cukup (sedang)
Lebih dari 0,70	Terlalu Mudah

¹⁵Zainal Arifin, *Evaluasi Pembelajaran*, (Jakarta : 2009), hal. 311

¹⁶Soenardi Djiwandono, *Tes Bahasa : Pegangan Bagi Pengajar Bahasa*, (Jakarta : Indeks , 2008), hal. 218-219

Contoh :

Penyebaran skor jawaban 10 orang peserta terhadap 10 butir item yang diajukan dalam tes hasil belajar bahasa Arab

Peserta	Skor yang dicapai oleh peserta untuk butir item									
	nomor									
	1	2	3	4	5	6	7	8	9	10
A	0	1	1	0	1	0	1	1	1	0
B	1	0	1	1	1	1	0	0	1	1
C	1	0	0	0	0	0	1	1	1	0
D	0	0	1	1	1	0	1	0	1	1
E	1	0	1	0	1	1	1	0	1	0
F	0	0	1	1	0	0	1	1	1	1
G	1	0	1	0	0	0	0	0	0	1
H	0	0	1	1	0	0	1	0	1	1
I	1	0	1	0	1	0	1	0	1	1
J	1	1	1	1	1	1	1	0	1	1
10	6	2	9	5	6	3	8	3	9	7

Maka dari contoh dapat dijelaskan tingkat kesukaran soalnya :¹⁷

1	$P = \frac{6}{10} = 0,60$ SEDANG	6	$P = \frac{3}{10} = 0,30$ SEDANG
2	DANG $P = \frac{2}{10} = 0,20$ TERLALU SUKAR	7	DANG $P = \frac{8}{10} = 0,80$ TERLALU MUDAH
3	RLA $P = \frac{9}{10} = 0,90$ TERLALU MUDAH	8	RLA $P = \frac{3}{10} = 0,30$ SEDANG
4	RLA $P = \frac{5}{10} = 0,50$ SEDANG	9	DANG $P = \frac{9}{10} = 0,90$ TERLALU MUDAH
5	DANG $P = \frac{6}{10} = 0,60$ SEDANG	10	RLA $P = \frac{7}{10} = 0,70$ SEDANG

Sedangkan rumus untuk menghitung tingkat kesukaran soal essay adalah :

¹⁷Anas Sudijono, *Pengantar Evaluasi Pendidikan*.....372-376

Mean
Jumlahskor Minimum

Contoh :

NO	NAMA SISWA	JUMLAH SOAL					TOTAL SKOR
		1	2	3	4	5	
1	A	6	6	12	12	6	42
2	B	6	6	6	12	12	42
3	C	6	6	12	6	6	36
4	D	12	6	12	12	12	54
5	E	6	6	12	12	12	48
6	F	6	6	12	12	12	48
7	G	12	6	12	12	6	48
8	H	6	6	6	6	12	36
9	I	12	12	12	6	12	54
10	J	6	6	6	12	6	36
11	K	6	6	6	6	6	30
12	L	6	6	6	12	6	36
13	M	6	6	6	12	12	42
14	N	6	6	12	12	12	48
15	O	6	6	6	12	12	42
16	P	6	12	6	6	6	36
17	Q	6	6	12	6	6	36
18	R	6	12	6	12	6	42
19	S	6	6	6	6	12	36
20	T	12	6	12	0	12	42
		144	138	180	186	186	

Penjelasan :

Skor soal essay yang ditetapkan adalah 60, tingkat kesukaran soal essay :

Rumus : $Mean = \frac{NP}{N}$

1. $Mean = \frac{144}{20} = 7,2$

Tingkat Kesukaran = $\frac{7,2}{60} = 0,12$ (sulit)

$$2. \text{ Mean} = \frac{138}{20} = 6,9$$

$$\text{Tingkat Kesukaran} = \frac{6,9}{60} = 0,115 = 0,12 \text{ (sulit)}$$

$$3. \text{ Mean} = \frac{180}{20} = 9$$

$$\text{Tingkat Kesukaran} = \frac{9}{60} = 0,15 \text{ (sulit)}$$

$$4. \text{ Mean} = \frac{186}{20} = 9,3$$

$$\text{Tingkat Kesukaran} = \frac{9,3}{60} = 0,155 = 0,16 \text{ (sulit)}$$

$$5. \text{ Mean} = \frac{186}{20} = 9,3$$

$$\text{Tingkat Kesukaran} = \frac{9,3}{60} = 0,155 = 0,16 \text{ (sulit)}^{18}$$

2. Daya Pembeda

Selain tingkat kesulitan, aspek lain dari butir tes yang dijadikan sasaran analisis adalah kemampuan butir tes untuk membedakan peserta tes yang mampu dan yang kurang mampu dalam menjawab pertanyaan tes atau mengerjakan tugas tes dengan benar. Kemampuan butir tes untuk membedakan peserta tes tersebut dikenal dengan daya pembeda.¹⁹ Daya pembeda soal berkaitan dengan kemampuan soal untuk mengetahui dan membedakan antara siswa yang pandai (menguasai materi) dan siswa yang kurang pandai (tidak/kurang menguasai materi).²⁰

Cara untuk mengetahui daya pembeda ini adalah dengan mengurutkan skor tertinggi paling atas sampai skor terendah, lalu dibagi 2, rumus yang digunakan adalah: $D = \frac{BA}{JA} - \frac{BB}{JB} = PA - PB$

Keterangan :

JA : Banyaknya peserta kelompok atas

JB : Banyaknya peserta kelompok bawah

BA : Banyak Peserta kelompok atas yang menjawab benar

BB : Banyak Peserta kelompok bawah yang menjawab benar

Pedoman yang digunakan untuk menginterpretasikan daya beda adalah:

a. Kurang dari 0,20 = Jelek

¹⁸Hasil Penelitian Lapangan Perkuliahan Evaluasi Pembelajaran Bahasa Arab S1

¹⁹Soenardi Djiwandono, *Tes Bahasa : Pegangan Bagi Pengajar Bahasa*,220

²⁰Indah Rahmi Nur Fauziah, dkk, 2020, *Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (Hots)*, LISANUNA, Vol. 10, No. 1

- b. 0,20-0,40 = Sedang
- c. 0,40-0,70 = Baik
- d. 0,70-1,00 = Baik Sekali
- e. Tanda negatif (-) = Jelek Sekali²¹

Cara menghitung daya beda soal essay : Daya Beda Essay = $\frac{\text{Mean KA} - \text{Mean KB}}{\text{Skor Maksimum Soal}}$

3. Fungsi Distraktor

Dalam tes objektif selalu digunakan alternatif jawaban yang mengandung dua unsur sekaligus, yaitu jawaban yang tepat dan jawaban yang salah sebagai penyesat(distraktor). Tujuannya adalah sebagai pengecoh bagi yang kurang mampu untuk dapat dibedakan dengan yang mampu. Jika sebuah distraktor dipilih oleh 5% lebih dari semua peserta tes, maka sebuah distraktor yang bersangkutan berfungsi, dan jika sebuah distraktor dipilih kurang dari 5% dari semua peserta tes maka distraktor yang bersangkutan tidak berfungsi.²² Adapun rumus dari fungsi distraktor ini adalah :

Rumus ; mencari P dan D = PA – PB

PENUTUP

Tes hasil belajar harus dapat mengukur secara jelas hasil belajar yang telah ditetapkan sesuai dengan tujuan. Butir-butir soal tes hasil belajar harus merupakan sampel yang representatif dari populasi bahan pelajaran yang telah diajarkan. Bentuk soal yang dikeluarkan dalam tes hasil belajar harus dibuat bervariasi. Tes hasil belajar harus di desain sesuai dengan kegunaannya untuk memperoleh hasil yang diinginkan. Tes hasil belajar harus memiliki reliabilitas yang dapat diandalkan. Tes hasil belajar disamping harus dapat dijadikan alat pengukur keberhasilan belajar siswa. Adapun Tahapan –tahapan yang dilalui dalam penyusunan tes bahasa, yaitu: Persiapan, Pemilihan materi tes, Menentukan bentuk dan jenis tes, Menentukan jumlah butir tes, Menentukan skor, Membuat kisi-kisi, Menyusun butir tes berdasarkan kisi-kisi, Uji coba tes yang telah disusun. Sedangkan cara menganalisis hasil tes

²¹Yelfi Dewi, *Evaluasi dan tes Bahasa Arab Pedoman bagi Guru, Mahasiswa dan Praktisi Pendidikan Bahasa Arab*.....hal. 104-105

²²Sri Suharti, 2017, *Kualitas Tes Bahasa Arab dan Prestasi Peserta Didik Madrasah Tsanawiyah Kabupaten Bantul (Analisis Butir Soal UAMBN Tahun Ajaran 2013/2014)*, Jurnal Pendidikan Madrasah, Volume 2, Nomor 1, Mei,, P-ISSN: 2527-4287-E-ISSN: 2527-6794

bahasa Arab :adalahdenganmenganalisis Tingkat kesukaransoal, DayaPembedadanfungsi distractor.

DAFTAR KEPUSTAKAAN

- Anas Sudijono, 2003, *Pengantar Evaluasi Pendidikan*, Jakarta : PT Raja Grafindo Persada.
- Erta Mahyudin, 2012, *Pengembangan Evaluasi dan Tes Bahasa Arab*, Tangerang: Alkitabiah.
- Fatimah Depi Susanty, 2016, Analisis Validasi Soal Tes Hasil Belajar Pada Pelaksanaan Pembelajaran Bahasa Arab Di Pusat Pengembangan Bahasa (P3b) Uin Suska Riau. Kutubkhanah: Jurnal Penelitian sosial keagamaan. Vol.19. No.2 Juli –Desember.
- Gito Supriadi, 2011, *Pengantar dan Teknik Evaluasi Pembelajaran*, Malang: Intimedia Press.
- Indah Rahmi Nur Fauziah, dkk. 2020, *Analisis Kualitas Tes Bahasa Arab Berbasis Higher Order Thinking Skill (Hots)*. LISANUNA. Vol. 10, No. 1
- M. Ainin, dkk. 2006, *Evaluasi Dalam Pembelajaran Bahasa Arab*, Malang: MISYKAT.
- Moh. Matsna HS., *Pengembangan Evaluasi dan Tes Bahasa Arab*, Tangerang: Alkitabiah.
- Soenardi Djiwandono, 2008, *Tes Bahasa: Pegangan Bagi Pengajar Bahasa*, Jakarta : Indeks.
- Sri Suharti, 2017, *Kualitas Tes Bahasa Arab dan Prestasi Peserta Didik Madrasah Tsanawiyah Kabupaten Bantul (Analisis Butir Soal UAMBN Tahun Ajaran 2013/2014)*, Jurnal Pendidikan Madrasah, Volume 2, Nomor 1, Mei, P-ISSN: 2527-4287, E-ISSN: 2527-6794.
- Yelfi Dewi, 2008, *Evaluasi dan Tes Bahasa Arab Pedoman bagi Guru, Mahasiswa dan Praktisi Pendidikan Bahasa Arab*, Bukittinggi : STAIN Bukittinggi Press.
- Zainal Arifin, 2009, *Evaluasi Pembelajaran*, Jakarta: tt.