

**ANALISIS VALIDASI SOAL TES HASIL BELAJAR PADA
PELAKSANAAN PEMBELAJARAN BAHASA ARAB
DI PUSAT PENGEMBANGAN BAHASA (P3B) UIN SUSKA RIAU**

¹Fatimah Depi Susanty

Universitas Islam Negeri Sultan Syarif Kasim Riau

Abstract

One of the main conditions that must be fulfilled related to the management quality system ISO 9001:2000 is the verification and calibration of the learning achievement test. This research aimed at finding the test type done by P3B Arabiclecturers, the steps done to develop and to validate the learning achievement test, and the characteristics of the learning achievement test quality which theoretically analyzed. This survey research included 192 lecturers as the population who taught basic science and technology subject matters in semester gasal 2014/2015. 92 of them was established by using proportional random sampling technique as the sample. Documentation and questionnaire technique were used to collect data. The data collected was analyzed by using descriptive quantitative method. The result of this research shows that 1) 55,42% P3B UIN Suska Riau Arabic lecturers write essay test for measuring learning achievement, 13,03% writes objective test and 31,55% combines between essay and objective test; 2) in the developing processes test: 35,87% lecturers make the blue-print of test and 23,92% do not make it, while only 13,59% lecturers who make the new test and 2,74% lecturers used the last year test and there are 46,19% lecturers combined the last year test and the new test; 3) 54,89% lecturers recheck test and 3,73% do not do it, 25,50% lecturers recheck by themselves and 17,85% lecturers recheck together in group of the same subject matter and fully rechecked by 2,18% others lecturers, for rechecking of the esei test done by 32,85% lecturers whereas the obyective test rechecked by 11,18%, the aspect that most rechecked was the relevance between the test and the learning competence which done by 91% lecturers; and 4) theoretically, 87% the test done by lecturers are good related to the essence of test quality.

Key words: *Achievement test, test quality and validation*

¹ Dosen Fakultas Tarbiyah dan Keguruan UIN Suska Riau

Abstrak

Salah satu syarat yang harus dipenuhi dengan dilaksanakannya sistem manajemen mutu ISO 9001-2000 adalah bahwa semua perangkat soal tes hasil belajar harus divalidasi sehingga instrumen tersebut dapat memenuhi kaidah-kaidah instrumen yang berkualitas untuk mengukur kompetensi hasil belajar. Penelitian ini bertujuan untuk mengungkap proses validasi soal ujian akhir semester yang dilakukan oleh dosen Bahasa Arab di P3B UIN SUSKA RIAU dan menganalisis secara teoretis (kualitatif) kualitas soal ujian tersebut. Penelitian ini menggunakan teknik survei dengan melibatkan populasi 192 orang dosen Bahasa Arab di P3B UIN SUSKA RIAU yang mengajar bidang studi pada semester gasal 2014/2015. Sampel sebanyak 92 orang sebagai sumber data dipilih secara random dan proporsional (proportional random sampling technique). Data dikumpulkan menggunakan teknik dokumentasi dan kuesioner. Data yang terkumpul dianalisis secara deskriptif kuantitatif. Hasil penelitian menunjukkan bahwa: 1) 55,42% dosen Bahasa Arab P3B UIN SUSKA RIAU membuat soal tes hasil belajar dalam bentuk esei, 13,03% membuat soal tes objektif dan 31,55% membuat soal dalam bentuk kombinasi tes esei dan objektif; 2) pada pengembangan soal ujian: 35,87% dosen membuat kisi-kisi soal dan 23,92% tidak membuat, sedangkan pada penulisan soal hanya 13,59% dosen membuat soal baru dan 2,74% dosen mengambil soal tahun lalu serta 46,19% kombinasi antara soal baru dan soal tahun lalu; 3) 54,89% dosen melakukan telaah ulang dan 3,73% tidak menelaah ulang, 25,50% melakukan telaah sendiri dan 17,85% ditelaah dengan dosen sebidang serta 2,18% dilakukan sepenuhnya oleh dosen lain, untuk tes esei telaah teoretis kualitas soal dilakukan oleh 32,85% dosen sedangkan untuk tes objektif hanya sebagian kecil yaitu 11,18% dosen, aspek yang banyak ditelaah adalah kesesuaian materi soal dengan kompetensi hasil belajar yang dilakukan oleh 91% dosen; dan 4) Secara teoretis, kualitas seluruh perangkat tes esei dan tes objektif yang dibuat dosen yang dinilai berdasarkan kaidah-kaidah soal yang baik mencapai sekitar 87%.

Kata kunci: *Tes Hasil Belajar, Kualitas Soal, dan Validasi.*

Pendahuluan

Mata kuliah Bahasa Arab merupakan mata kuliah wajib yang harus ditempuh dan dipelajari oleh para mahasiswa seluruh fakultas di UIN SUSKA RIAU dari semester I hingga semester III. Pembelajaran Bahasa Arab ini dilaksanakan secara berturut-turut di pusat pengembangan bahasa. Karena mata kuliah ini merupakan mata kuliah wajib, maka mahasiswa tidak punya pilihan lain, tidak memandang asal sekolah mahasiswa tersebut. Sehingga dalam proses pembelajaran banyak sekali ditemukan kendala-kendala yang sangat signifikan. Namun demikian, UIN SUSKA RIAU telah menyiapkan tenaga-tenaga pengajar yang profesional di bidangnya.

Para tenaga pengajar tidak hanya bertanggung jawab pada proses transaksi ilmu, namun lebih dari itu tenaga pengajar dalam hal ini dosen dituntut juga untuk membuat instrumen penilaian dan pengukuran untuk mengetahui sejauhmana para

mahasiswa mampu dan dapat menguasai materi Bahasa Arab tersebut dalam sebuah proses evaluasi atau lebih tepatnya tes. Tes yang digunakan untuk mengukur kemampuan para mahasiswa di pusat bahasa ada dua bentuk tes, yaitu tes lisan dan tes tertulis. Dalam usulan penelitian ini peneliti menfokuskan pada tes tertulis karena dalam pembuatan instrument tes tertulis dosen mempunyai peranan yang signifikan, yaitu dosen diwajibkan membuat soal yang berhubungan dengan tema yang sudah ditentukan oleh lembaga pusat bahasa. Bentuk soal yang dibuat oleh dosen adalah objektif dengan 50 butir soal. Dengan demikian, akan banyak soal yang terkumpulkan di pusat bahasa. Soal-soal ini selanjutnya akan disaring sehingga nanti akan terpilihlah butir soal yang layak atau berkualitas dimuat dalam proses evaluasi hasil.

Butir soal dikatakan berkualitas apabila butir soal yang dibuat memiliki validitas yang tinggi. Validitas

merupakan suatu alat penilaian terhadap ketepatan soal agar seorang dosen benar-benar menilai apa yang ingin di nilai. Menurut Uno (2006: 103), validitas butir soal ada empat macam yang berasal dan dasar pembagian jenis, yaitu (a) validitas logis meliputi validitas isi dan validitas konstruksi, (b) validitas empiris meliputi validitas ada sekarang dan validitas prediksi.

Selain ke empat macam validitas soal di atas, masih ada satu jenis validitas yang dianggap mempunyai peranan penting dalam menentukan kualitas butir soal, yaitu validitas item tes hasil belajar. Validitas item dari suatu tes adalah ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian yang tidak terpisahkan dari tes sebagai suatu totalitas) dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.

Apabila kita mau memperhatikan secara cermat, maka tes-tes hasil belajar yang dibuat atau disusun oleh para pengajar, baik guru, dosen, dan staf pengajar lainnya,

sebenarnya adalah merupakan kumpulan dari sekian banyak butir-butir item, dengan item mana para penyusun tes ingin mengukur dan mengungkap hasil belajar yang telah dicapai oleh masing-masing individu peserta didik setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Pernyataan ini menegaskan bahwa sebenarnya setiap butir item yang ada dalam tes hasil belajar itu adalah merupakan bagian yang tidak terpisahkan dari tes hasil belajar tersebut sebagai suatu totalitas.

Permasalahan validitas item yang terlihat jelas adalah ketika para guru, dosen atau staf pengajar hanya fokus pada pembuatan butir soal saja tanpa mempertimbangkan hal-hal yang berhubungan dengan validitas butir soal. Secara sistematis pembuatan butir soal, seharusnya setelah penulisan butir soal diadakan *try out* atau uji coba untuk mengetahui sejauhmana butir soal yang dibuat oleh para guru dan dosen itu masuk kategori valid. Butir soal yang valid lah yang layak untuk di jadikan bahan dalam pelaksanaan tes hasil belajar.

Seperti telah dikemukakan di muka bahwa tugas utama dosen adalah melaksanakan pembelajaran. Untuk mengetahui hasil pembelajarannya, dosen yang bersangkutan harus melakukan pengukuran dengan membuat tes hasil belajar. Tes buatan dosen menjadi sangat penting artinya dalam kaitannya dengan efektivitas proses pembelajaran dan peningkatan hasil belajar mahasiswa. Berdasarkan uraian di atas maka permasalahan yang akan diteliti oleh peneliti adalah “bagaimana analisis validitas ítem tes hasil belajar yang dibuat oleh para dosen bahasa arab di Pusat Pengembangan Bahasa UIN SUSKA RIAU”.

Pembelajaran Bahasa Arab

Pengertian Pembelajaran adalah upaya untuk belajar. Kegiatan ini akan mengakibatkan siswa mempelajari sesuatu dengan cara efektif dan efisien (Muhaimin, dkk., 1996: 99). Sebagaimana hal yang disebutkan oleh Nababan bahwasannya arti pembelajaran adalah

nominalisasi proses untuk membelajarkan (Jos D. Parera, 1997: 24-25). Seharusnya pembelajaran bermakna “proses membuat atau menyebabkan orang lain belajar. Menurut Oemar Hamalik (1995: 57), Pembelajaran adalah suatu kombinasi yang tersusun meliputi unsur-unsur manusiawi, material, fasilitas, perlengkapan, dan prosedur yang saling mempengaruhi untuk mencapai tujuan pembelajaran, dalam hal ini manusia terlibat dalam sistem pengajaran terdiri dari siswa, guru, dan tenaga lainnya, materi meliputi; buku-buku, papan tulis, dan lain-lain. Fasilitas dan perlengkapan terdiri dari ruang kelas dan audiovisual. Prosedur meliputi jadwal dan metode penyampaian informasi, praktik belajar, ujian, dan sebagainya. Pembelajaran disebut juga sebagai proses perilaku dengan arah positif untuk memecahkan masalah personal, ekonomi, sosial, dan politik yang ditemui oleh individu, kelompok, dan komunitas.

Mata pelajaran Bahasa Arab merupakan suatu mata pelajaran yang

diarahkan untuk mendorong, membimbing, mengembangkan, dan membina kemampuan serta menumbuhkan sikap positif terhadap Bahasa Arab, baik reseptif maupun produktif. Kemampuan reseptif yaitu kemampuan untuk memahami pembicaraan orang lain dan memahami bacaan. Kemampuan produktif yaitu kemampuan menggunakan bahasa sebagai alat komunikasi, baik secara lisan maupun tulis. Kemampuan berbahasa Arab serta sikap positif terhadap Bahasa Arab tersebut sangat penting dalam membantu memahami sumber ajaran Islam, yaitu al-Qur'an dan Hadits, serta kitab-kitab Bahasa Arab yang berkenaan dengan Islam bagi peserta didik.

Untuk itu Bahasa Arab di tingkat Universitas dipersiapkan untuk pencapaian kompetensi dasar berbahasa, yang mencakup empat keterampilan berbahasa yang diajarkan secara integral, yaitu menyimak, berbicara, membaca, dan menulis. Meskipun begitu, pada tingkat pendidikan dasar (*elementary*)

dititikberatkan pada kecakapan menyimak dan berbicara sebagai landasan berbahasa. Pada tingkat pendidikan menengah (*intermediate*), keempat kecakapan berbahasa diajarkan secara seimbang. Adapun pada tingkat pendidikan lanjut (*advanced*) dikonsentrasikan pada kecakapan membaca dan menulis, sehingga peserta didik diharapkan mampu mengakses berbagai referensi Bahasa Arab.

Mata pelajaran Bahasa Arab memiliki tujuan sebagai berikut:

- a. Mengembangkan kemampuan berkomunikasi dalam Bahasa Arab, baik lisan maupun tulis, yang mencakup empat kecakapan berbahasa.
- b. Menumbuhkan kesadaran tentang pentingnya Bahasa Arab sebagai salah satu bahasa asing untuk menjadi alat utama belajar, khususnya dalam mengkaji sumber-sumber ajaran Islam.
- c. Mengembangkan pemahaman tentang saling keterkaitan antara bahasa dan budaya serta memperluas cakrawala budaya dan

melibatkan diri dalam keragaman budaya.

Tes Hasil Belajar

Kata tes berasal dari bahasa Prancis kuno yang berarti piring untuk menyisihkan logam-logam mulia. Maksudnya adalah dengan menggunakan alat berupa piring akan dapat diperoleh jenis-jenis logam mulia yang bernilai tinggi (Brown and A. Page, 1970: 7). Dalam perkembangannya dan seiring kemajuan zaman, tes berarti ujian atau percobaan. Ada beberapa istilah yang memerlukan penjelasan sehubungan dengan uraian di atas yaitu tes, testing, tester, dan testee, yang masing-masing mempunyai pengertian berbeda namun erat kaitannya dengan tes.

1. Tes adalah alat atau prosedur yang dipergunakan dalam rangka pengukuran dan penilaian;
2. Testing berarti saat dilaksanakannya pengukuran dan penilaian atau saat pengambilan tes;

3. Tester artinya orang yang melaksanakan tes atau orang yang disertai untuk melaksanakan pengambilan tes terhadap para responden;
4. Testee adalah pihak yang sedang dikenai tes.

Ada beberapa pendapat dari beberapa ahli tentang pengertian tes. Menurut Anne Anastasi dalam karya tulisnya yang berjudul *Psychological Testing*, yang dimaksud dengan tes adalah alat pengukur yang mempunyai standar yang objektif sehingga dapat digunakan secara meluas, serta dapat digunakan sebagai cara untuk mengukur dan membandingkan keadaan psikis atau tingkah laku individu. Menurut Lee J. Cronbach (1990: 12), tes merupakan suatu prosedur yang sistematis untuk membandingkan tingkah laku dua orang atau lebih. Sedangkan menurut Goodenough, seperti yang dikutip Robert Lado (1961: 10), tes adalah suatu tugas atau serangkaian tugas yang diberikan kepada individu atau kelompok individu, yang dimaksud

untuk membandingkan kecakapan satu sama lain.

Dari pengertian dari para ahli tersebut dalam dunia pendidikan dapat disimpulkan bahwa tes adalah cara yang digunakan atau prosedur yang ditempuh dalam rangka pengukuran dan penilaian di bidang pendidikan, yang memberikan tugas dan serangkaian tugas yang diberikan oleh guru sehingga dapat dihasilkan nilai yang melambangkan tingkat laku atau prestasi peserta didik.

Jenis dan Bentuk Tes Hasil Belajar

Dari segi bentuk soal dan kemungkinan jawabannya tes dibagi menjadi 2 bagian yakni:

1. Tes Essay (uraian)

Tes essay adalah tes yang disusun dalam bentuk pertanyaan terstruktur dan mahasiswa menyusun, mengorganisasikan sendiri jawaban tiap pertanyaan itu dengan bahasa sendiri. Tes essay ini sangat bermanfaat untuk mengembangkan kemampuan dalam menjelaskan atau

mengungkapkan suatu pendapat dalam bahasa sendiri. Subino, (1987: 2) menyatakan bahwa berdasarkan tingkat kebebasan jawaban yang dimungkinkan dalam tes bentuk uraian, butir-butir soal dalam ini dapat dibedakan atas butir-butir soal yang menuntut jawaban bebas. Butir-butir soal dengan jawaban terikat cenderung akan membatasi, baik isi maupun bentuk jawaban; sedangkan butir soal dengan jawaban bebas cenderung tidak membatasi, baik isi maupun jawaban.

2. Tes Objektif

Tes objektif adalah tes yang disusun sedemikian rupa dan telah disediakan alternatif jawabannya. Tes ini terdiri dari berbagai macam bentuk, antara lain:

- Tes Betul-Salah (True False)
- Tes Pilihan Ganda (Multiple Choice)
- Tes Menjodohkan (Matching)
- Tes Analisa Hubungan (Relationship Analysis)

Pada prinsipnya, bentuk tes objektif di atas mempunyai kelemahan dan kebaikannya, akan tetapi biasanya bentuk objektif dapat menteskan semua bahan yang telah diajarkan, sedangkan bentuk uraian agak sukar untuk mengukur semua bahan yang sudah diajarkan, karena ruang lingkup bentuk tes tersebut sangat sempit. Untuk lebih jelasnya perlu diterangkan dahulu kelemahan dan kebaikan tes bentuk objektif. Keuntungan atau kebaikan bentuk objektif dalam evaluasi hasil belajar bahasa Indonesia bagi siswa adalah tes bentuk objektif; (1) tepat untuk mengungkapkan hasil belajar yang bertatanan pengetahuan, pemahaman, aplikasi, dan analisis, (2) mempunyai dampak belajar yang mendorong mahasiswa untuk mengingat, menafsirkan, dan menganalisis pendapat, dan (3) jawaban yang diberikan dapat menggambarkan ranah tujuan pendidikan menurut Bloom, khususnya ranah *cognitive domain*. Sedangkan kelemahannya bahwa tes objektif; (1) mahasiswa tidak dituntut untuk mengorganisasikan jawaban, karena

jawabannya sudah disediakan, (2) mahasiswa ada kemungkinan dapat menebak jawaban yang telah tersedia, (3) tidak dapat mengungkap proses berpikir dan bernalar, dan (4) hanya mengukur ranah kognitif yang paling rendah tidak mengungkap kemampuan yang lebih kompleks. Hal ini sebagaimana yang diungkapkan Gronlund (1985: 36) bahwa *...objective test items can be used to measure a variety of knowledge outcomes ...the most generally useful is the multiple choice items...but other items types also have a place*. Pernyataan tersebut menunjukkan bahwa item-item tes objektif dapat digunakan untuk mengukur berbagai hasil belajar yang berupa pengetahuan. Umumnya yang paling berguna adalah item bentuk pilihan jamak, sementara itu, tipe item objektif yang lainnya punya peran tersendiri.

Ciri-ciri tes yang baik

Menurut Arikonto (1992), sebuah tes yang dapat dikatakan baik sebagai alat pengukur harus memiliki persyaratan tes, yaitu memiliki:

1. Validitas

Sebuah tes disebut valid 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.

2. Reliabilitas

Berasal dari kata *reliable* yang artinya dapat dipercaya. Tes dapat dikatakan dapat dipercaya jika memberikan hasil yang tetap apabila diteskan berkali-kali. Sebuah tes dikatakan *reliabel* apabila hasil-hasil tes tersebut menunjukkan ketetapan. Jika dihubungkan dengan validitas, maka: validitas adalah ketepatan dan reliabilitas adalah ketetapan.

3. Objektivitas

Sebuah tes dikatakan memiliki objektivitas apabila dalam melaksanakan tes itu tidak ada faktor subjektif yang mempengaruhi. hal ini terutama terjadi pada sistem *scoringnya*. Apabila dikaitkan dengan reliabilitas, maka objektivitas menekankan ketetapan pada sistem *scoringnya*, sedangkan reliabilitas menekankan ketetapan dalam hasil tes.

4. Praktisibilitas

Sebuah tes dikatakan memiliki praktisibilitas yang tinggi apabila tes tersebut bersifat praktis dan mudah pengadministrasiannya. Tes yang baik adalah yang mudah dilaksanakan, mudah pemeriksaannya, dan dilengkapi dengan petunjuk-petunjuk yang jelas.

5. Ekonomis

Yang dimaksud ekonomis di sini ialah bahwa pelaksanaan tes tersebut tidak membutuhkan ongkos atau biaya yang mahal, tenaga yang banyak, dan waktu yang lama.

Pengertian Validitas Item

Dimaksud dengan validitas item dari suatu tes adalah ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian tak terpisahkan dari tes sebagai sesuatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.

Apabila kita mau memperhatikan secara cermat, maka tes-tes hasil belajar yang dibuat atau disusun oleh para pengajar, baik guru, dosen, dan staf pengajar lainnya, sebenarnya merupakan kumpulan dari sekian banyak butir-butir item; dengan item mana para penyusun tes ingin mengukur atau mengungkap hasil belajar yang telah dicapai oleh masing-masing individu peserta didik, setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu. Pernyataan itu mengandung makna, bahwa sebenarnya setiap butir item yang ada dalam tes hasil belajar itu merupakan bagian tak terpisahkan dari tes hasil belajar tersebut sebagai suatu totalitas.

Eratnya hubungan antara butir item dengan tes hasil belajar sebagai

suatu totalitas itu kiranya dapat dipahami dari kenyataan, bahwa semakin banyak butir-butir item yang dapat dijawab dengan betul oleh testee, maka skor-skor total hasil tes tersebut akan semakin tinggi. Sebaliknya semakin sedikit butir-butir item yang dapat dijawab dengan betul oleh testee, maka skor-skor total hasil tes itu akan semakin rendah dan semakin menurun.

Pernyataan tersebut merupakan petunjuk bahwa semakin besar “dukungan” yang diberikan oleh butir-butir item (sebagai bagian tak terpisahkan dari tes) terhadap tes hasil belajar (sebagai suatu totalitas), maka tes tersebut akan semakin dapat menunjukkan “kemantapannya”. Sebaliknya, semakin kecil “dukungan” yang diberikan oleh masing-masing butir item terhadap tes sebagai suatu totalitas, maka tes menjadi semakin “kurang mantap”.

Apabila pernyataan tersebut kita kaitkan dengan validitas item yang sedang kita bicarakan ini, maka dapat dipahami bahwa sebenarnya validitas tes itu akan sangat dipengaruhi oleh atau sangat tergantung pada validitas

yang dimiliki oleh masing-masing butir item yang membangun tes tersebut. Makna yang terkandung dalam pernyataan itu lebih lanjut adalah, bahwa validitas dari masing-masing butir item yang membangun tes itu, akan dapat diketahui dengan jalan melihat besar kecilnya dukungan yang diberikan oleh masing-masing butir item yang bersangkutan terhadap tes sebagai keseluruhan.

Persoalan tentang validitas item itu kiranya tidak terlalu mendesak untuk ditangani, kalau saja berdasarkan uji validitas ternyata bahwa tes hasil belajar yang dibuat oleh guru, dosen atau staf pengajar lainnya ternyata telah memiliki validitas tes yang tinggi, sehingga tes hasil belajar itu sebagai totalitas sudah dapat dikatakan handal dan tidak perlu diragukan ketepatan mengukurnya. Namun persoalan akan segera muncul, jika setelah dilakukan uji validitas tes terhadap tes yang dijadikan alat pengukur itu ternyata menghasilkan kesimpulan bahwa tes hasil belajar itu validitasnya sangat rendah, sehingga

dapat dimasukkan dalam kategori tes hasil belajar yang invalid.

Munculnya persoalan yang disebutkan terakhir itu sudah seharusnya mendapatkanantisipasi atau respon dari penyusun tes. Para penyusun tes harus cukup tanggap, bahwa rendahnya validitas tes itu merupakan salah satu indikator bahwa dalam tes hasil belajar yang mereka susun itu “telah terjadi hal-hal yang kurang beres” sehingga mereka perlu melakukan tindakan “mawas diri”.

Dalam rangka “mawas diri” itu, penyusun tes hasil belajar perlu melakukan penelusuran kembali terhadap tes hasil belajar bagi peserta didik, yang ternyata tidak memiliki daya ketepatan mengukur seperti yang diharapkan (Sutardi Wirasasmita, 1998). Penelusuran atau pelacakan kembali itu dilakukan oleh penyusun tes, guna mengetahui butir-butir item mana sajakah dari keseluruhan butir item yang telah dikeluarkan dalam tes hasil belajar, yang telah menyebabkan rendahnya validitas tes tersebut sebagai suatu totalitas.

Teknik Pengujian Validitas Item Hasil Belajar

Dari uraian yang telah dikemukakan di atas, kiranya menjadi cukup jelas bahwa sebutir item dapat dikatakan telah memiliki validitas yang tinggi atau dapat dinyatakan valid, jika skor-skor pada butir item yang bersangkutan memiliki kesesuaian atau kesejajaran arah dengan skor totalnya; atau dengan bahasa statistik: ada korelasi positif yang signifikan antara skor item dengan skor totalnya. Skor total di sini berkedudukan sebagai variabel terikat (*dependent variable*), sedangkan skor item berkedudukan sebagai variabel bebasnya (*independent variabel*). Kalau demikian, maka untuk sampai pada kesimpulan bahwa item-item yang ingin diketahui validitasnya, yaitu valid atukah tidak, kita dapat menggunakan teknik korelasi sebagai teknik analisisnya. Sebutir item dapat dinyatakan valid, apabila skor item yang bersangkutan terbukti mempunyai korelasi positif yang signifikan dengan skor totalnya.

Persoalan berikutnya ialah memilih dan menentukan jenis teknik korelasi yang dipandang tepat untuk digunakan dalam rangka uji validitas item itu. Seperti diketahui, pada tes objektif hanya ada dua kemungkinan jawaban, yaitu betul dan salah. Setiap butir soal yang dijawab dengan betul umumnya diberi skor 1 (satu), sedangkan untuk setiap jawaban salah diberikan skor 0 (nol). Jenis data seperti ini, yaitu: betul-salah, ya-tidak atau yang sejenis dengan itu, dalam dunia ilmu statistik dikenal dengan nama data diskret murni atau data dikotomik. Sedangkan skor total yang dimiliki oleh masing-masing individu testee adalah merupakan hasil penjumlahan dari setiap skor yang dimiliki oleh masing-masing butir item (misalnya:

$0+1+1+0+1+0+1+1+0+0+1=6$) itu merupakan data kontinyu.

Menurut teori yang ada, apabila variabel I berupa data diskret murni atau data dikotomik, sedangkan variabel II berupa data kontinyu, maka teknik korelasi antara variabel I dengan variabel II itu adalah teknik

korelasi point biserial, di mana angka indeks korelasi yang diberi lambang r_{pbi} dapat diperoleh dengan menggunakan rumus (Suharsimi Arikunto, 1992: 180):

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

Di mana:

r_{pbi} = Koefisien korelasi point biserial yang melambangkan kekuatan korelasi antara variabel I dengan variabel II, yang dalam hal ini dianggap sebagai Koefisien Validitas Item.

M_p = Skor rata-rata hitung yang dimiliki oleh testee, yang untuk butir item yang bersangkutan telah dijawab dengan betul.

M_t = Skor rata-rata dari skor total.

SD_t = Deviasa standar dari skor total.

p = Proporsi testee yang menjawab betul terhadap butir item yang sedang diuji validitas itemnya.

q = Proporsi testee yang menjawab salah terhadap butir item yang sedang diuji validitas itemnya.

Metodologi Penelitian

A. Jenis Penelitian

Penelitian ini merupakan penelitian evaluatif yang mengevaluasi butir soal dengan menggunakan pendekatan kuantitatif. Penelitian evaluatif adalah jenis penelitian yang digunakan untuk menilai keberhasilan, manfaat, kegunaan, sumbangan, dan kelayakan suatu program, produk atau kegiatan suatu lembaga berdasarkan kriteria tertentu.

B. Lokasi dan Kondisi Penelitian

Lokasi penelitian ini adalah di Pusat Pengembangan Bahasa UIN SUSKA RIAU.

C. Populasi dan Sampel

Adapun populasi penelitian ini adalah seluruh dosen Bahasa Arab di Pusat Bahasa Sebanyak 192 orang, kemudian peneliti mengambil sampel secara acak sebanyak 92 orang dosen.

D. Teknik Pengumpulan Data

Teknik pengumpulan data adalah dengan dokumentasi, yaitu kumpulan soal yang ada di Pusat Pengembangan Bahasa UIN SUSKA RIAU. Soal inilah yang akan diuji

cobakan guna untuk mengetahui tingkat validitas butir soal.

E. Teknik Analisis Data

Teknik analisis data yang digunakan untuk menyelesaikan penelitian ini adalah menggunakan analisis deskriptif kuantitatif, yaitu dengan memaparkan dan menyajikan data yang ditemukan di lapangan dengan tampilan kuantitatif. Hal ini dikarenakan data yang ditemukan dan diselesaikan berupa angka atau hasil *try out* atau uji coba soal-soal bahasa arab yang ada di Pusat Pengembangan Bahasa.

Hasil Penelitian dan Pembahasan

Analisis data hasil penelitian difokuskan pada bentuk soal yang dibuat oleh dosen, langkah-langkah dalam penyusunan soal, langkah-langkah dalam memvalidasi soal, dan kualitas soal berdasarkan tinjauan teoretis. Deskripsi terhadap hasil analisis tersebut adalah seperti berikut.

1. Berdasarkan hasil analisis tampak bahwa lebih dari separoh dosen

Pusat Bahasa membuat soal tes hasil belajar dalam bentuk objektif, yaitu sebanyak 70%, sedangkan sekitar 30% membuat tes bentuk uraian. Penyusunan soal tes hasil belajar yang cenderung berbentuk objektif dilakukan mengingat sifat mata kuliah Bahasa Arab banyak yang bersifat aplikatif.

2. Pengembangan kisi-kisi, berdasarkan hasil analisis dapat dikemukakan bahwa sekitar 40,21% dosen Bahasa Arab di Pusat Bahasa yang memperhatikan kisi-kisi dalam proses penyusunan soal tes hasil belajar, sedangkan 35% dosen tidak membuat kisi-kisi, dan 23,92% dosen tidak memberikan tanggapan apakah membuat kisi-kisi atau tidak dengan alasan bahwa pada proses pembuatan sudah memperhatikan kompetensi hasil belajar yang ditetapkan.
3. Penulisan soal, berdasarkan deskripsi data tampak bahwa hanya sekitar 13,59% dosen yang membuat soal baru untuk ujian akhir semester mahasiswa,

sedangkan 46,19% dosen membuat soal kombinasi antara soal tahun lalu dan soal baru dan hanya sebagian kecil yang menggunakan soal tahun lalu. Di samping itu, terdapat 37,48% dosen tidak memberikan tanggapan terhadap pembuatan soal.

4. Validitas teoretis, hasil analisis hanya 54,89% dosen yang melakukan telaah ulang terhadap soal yang sudah disusun dalam rangka untuk menyiapkan soal tes hasil belajar yang berkualitas. Sisanya, 45,11% dosen tidak melakukan telaah ulang dengan alasan bahwa telaah ini telah dilakukan pada saat pembuatan soal. Semua aspek yang berkaitan dengan kualitas soal bentuk objektif sudah dilakukan telaah ulang secara teoretis oleh dosen (85,72%).
5. Kualitas soal, berdasarkan tinjauan teoretis deskripsi data dapat dikemukakan bahwa dengan telah dilakukannya telaah ulang secara teoretis terhadap aspek-aspek yang berkaitan dengan kualitas soal,

maka dapat dikatakan bahwa secara keseluruhan soal tes hasil belajar yang dibuat oleh dosen 85% baik. Namun demikian, hasil analisis yang dilakukan oleh peneliti masih menemukan adanya beberapa dapat dikemukakan bahwa 36% dari semua soal bentuk objektif tidak mencantumkan bobot nilai masing-masing butir, 28% soal memberi kebebasan memilih soal yang dikerjakan, dan 18% perangkat soal yang dinilai lemah pada gambar yang menyertai soal. Selain itu, masih ditemukan dalam 35% perangkat soal yang beberapa pada butir soal pokok soalnya tidak satu halaman dengan pilihan jawaban, 15% soal masih memiliki gambar yang menyertai soal termasuk tidak jelas. Di samping itu, terdapat 34% perangkat soal yang hanya menilai kompetensi hasil belajar terbatas pada ranah pengetahuan dan pemahaman.

Berdasarkan uraian singkat terhadap deskripsi data di muka dapat

diungkapkan empat hal pokok yang menjadi permasalahan dalam penelitian ini. *Pertama*, lebih dari separoh dosen mata kuliah Bahasa Arab di Pusat Pengembangan Bahasa P3B UIN Suska (70,00%) membuat tes hasil belajar dalam bentuk tes objektif bentuk pilihan ganda. Hal ini dapat dimengerti karena karakteristik dari matakuliah Bahasa Arab lebih cenderung pada aplikasi dan pemecahan masalah. Dengan bentuk tes objektif diharapkan mahasiswa dapat mengembangkan kemampuan penalarannya dalam menyelesaikan berbagai problem kebahasaan dengan praktis. Dengan demikian, tidak mengherankan jika hanya 30,00% dosen membuat soal tes hasil belajar dalam bentuk uraian. Menurut hemat penulis sebetulnya akan lebih bermakna lagi jika tes hasil belajarnya dibuat dengan tes objektif.

Kedua, dalam mengembangkan dan membuat soal, hanya 40,21% dosen membuat kisi-kisi soal, 23,92% tidak membuat kisi-kisi soal dan bahkan 35,87% dosen tidak jelas apakah membuat kisi-kisi soal atau

tidak. Melihat data tersebut dapat juga dikatakan bahwa dalam merencanakan dan mengembangkan soal tes hasil belajar para dosen belum seluruhnya melakukan langkah-langkah awal pembuatan tes sebagaimana yang diharapkan. Hal ini tentunya belum optimal karena untuk dapat membuat soal yang baik maka kisi-kisi soal sangat diperlukan. Dengan adanya kisi-kisi soal, maka soal yang dibuat diharapkan tidak menyimpang dari kompetensi dasar yang sudah ditetapkan. Di samping itu, sekitar 46,19% dari dosen mata kuliah Bahasa Arab di Pusat Bahasa UIN Suska membuat soal dengan cara kombinasi soal baru dan soal tahun lalu. Akan lebih baik jika semua soal tes dibuat baru karena sangat dimungkinkan soal tahun lalu masih mudah diingat oleh mahasiswa-mahasiswa yang kebetulan mengulang, kecuali ada modifikasi terhadap soal tahun tersebut. Berdasarkan analisis data hanya 13,59% dosen mata kuliah P3B UIN Suska yang membuat soal yang betul-betul baru. Persentase ini memang dirasa masih rendah karena semestinya

semua dosen dapat mengembangkan soal-soal tes yang baru sesuai dengan tuntutan kompetensi dan perkembangan ilmu.

Ketiga, dalam proses validasi dan verifikasi terhadap kualitas soal dapat dikatakan sudah cukup baik. Sebanyak 54,89% dosen P3B UIN Suska melakukan telaah ulang terhadap soal tes hasil belajar sebelum diujikan. Namun demikian, akan sangat baik dan berarti jika semua dosen melakukan telaah ulang terhadap soal-soal tes yang sudah disusun. Telaah ulang mencakup tiga hal yaitu dari aspek materi, konstruksi, dan bahasa. Dikarenakan hampir separoh dosen tidak melakukan telaah ulang secara teoretis, maka dapat dimengerti jika masih ada sebagian perangkat soal tes hasil belajar yang dibuat oleh dosen belum memenuhi kaidah-kaidah sebagai tes yang baik. Di samping itu, telaah ulang yang baik adalah jika dilakukan bersama-sama dengan dosen lain yang sebidang, tidak dilakukan sendiri. Berdasarkan analisis data, baru sebagian kecil telaah ulang dilakukan dengan dosen lain yang

sebidang, yaitu sekitar 17,85%. Padahal jika telaah ulang semua dilakukan dengan dosen lain yang sebidang maka ketidaktepatan bahasa dan substansi tes dapat diatasi karena adanya masukan dan saran dari berbagai pihak yang terlibat dalam telaah ulang tersebut.

Dalam menelaah soal-soal yang sudah dibuat dapat dikatakan sudah dilakukan terhadap semua aspek-aspek yang seharusnya dicermati. Namun demikian, masih ada beberapa aspek yang kurang diperhatikan oleh dosen. Misalnya, untuk tes objektif bentuk pilihan ganda, mencermati kejelasan gambar hanya dilakukan oleh 49,50% dosen, dan mencermati ketepatan bahasa dilakukan oleh 26,50% dosen. Hal ini dapat dikatakan bahwa dalam menelaah kedua aspek tersebut masih kurang tuntas. Padahal jika gambar yang menyertai butir soal kurang jelas akan mempengaruhi ketepatan mahasiswa dalam menjawab soal tes. Termasuk juga dari segi bahasa seharusnya tidak menimbulkan multitafsir atau beragam interpretasi.

Demikian juga telaah terhadap tes objektif, ternyata tidak ada separoh dari dosen P3B UIN Suska yang melakukan telaah secara teoretis terhadap semua aspek yang semestinya diperhatikan. Terdapat tiga aspek yang masih sangat kurang diperhatikan oleh dosen dalam membuat tes bentuk objektif, yaitu homogenitas pilihan jawaban (17,75%), panjang alternatif pilihan jawaban yang relatif sama (16,75%), dan telaah keberfungsian gambar/grafik (14%). Berdasarkan data juga dapat dikatakan bahwa telaah teoretis terhadap soal tes bentuk esei lebih diperhatikan dari pada telaah terhadap soal tes bentuk objektif.

Keempat, berdasarkan hasil analisis peneliti terhadap dokumen soal yang dijadikan sampel ditemukan beberapa kelemahan. Kelemahan yang paling dominan adalah bahwa soal tes lebih cenderung mengukur aspek pengetahuan dan pemahaman, belum banyak merambah pada aspek aplikasi, analisis, sintesis bahkan evaluasi. Padahal untuk tingkat mahasiswa pengukuran kemampuan belajar mulai dari aplikasi sampai dengan evaluasi

merupakan hal yang harus dan utama dilakukan. Di samping itu, untuk tes objektif masih ditemukan gambar-gambar yang kurang jelas, kebebasan memilih soal yang dikerjakan, dan bobot nilai untuk masing-masing butir. Hal ini semua akan mempengaruhi ketepatan kesimpulan hasil penilaian terhadap tingkat kompetensi belajar. Di samping itu, tes objektif, adanya ketidakhomogenan pilihan jawaban (panjang pendeknya pernyataan semua pilihan jawaban), dan terpisahnya halaman pokok soal dengan halaman jawaban menunjukkan bahwa telaah belum dilaksanakan secara mendalam. Dengan kata lain, ada hubungan antara kegiatan yang dilakukan dalam proses telaah kualitas soal dengan berbagai kelemahan soal yang digunakan untuk ujian.

Kesimpulan

Kecenderungan dosen P3B UIN Suska membuat soal tes hasil belajar dalam bentuk objektif didasarkan atas karakteristik mata kuliah yang bersifat praktis. Dengan bentuk tes objektif, maka lebih mudah dikembangkan soal-

soal yang mampu mengungkap kemampuan aplikatif, analisis, sintesis, dan evaluasi mahasiswa dari pada tes objektif.

Dalam pengembangan soal tes hasil belajar belum semua dosen (sekitar 76%) membuat kisi-kisi soal yang dijadikan acuan penulisan butir soal. Padahal kisi-kisi sangat penting artinya pada proses penulisan soal yang baik. Di samping itu, dalam penulisan soal, sebagian besar dosen melakukan kombinasi antara soal tahun lalu dan soal baru. Pembuatan soal baru akan lebih bermakna berkaitan dengan perkembangan sains dan teknologi.

Belum semua dosen mata kuliah bahasa arab di P3B UIN Suska melakukan telaah ulang secara teoretis menyangkut kualitas soal yang sudah dibuat. Meskipun demikian, ada satu hal yang cukup baik dan positif dalam telaah ulang yang sudah dilakukan oleh sebagian besar dosen, yaitu aspek yang paling banyak ditelaah ulang adalah kesesuaian materi soal dengan kompetensi hasil belajar.

Masih ditemukan beberapa kelemahan pada sejumlah perangkat soal, yaitu untuk tes objektif, aspek yang masih lemah adalah adanya ketidakseragaman panjang pendeknya pernyataan antar pilihan jawaban, terpisahnya halaman pokok soal dengan halaman jawaban. Kelemahan lain untuk bentuk tes tersebut adalah masih ada gambar-gambar yang menyertai soal yang kurang jelas.

Daftar Kepustakaan

- Brown, A. L , and Page. A. (1970). *Elements of Functional Analysis*. Butler & Tanner Ltd. Frome and London.
- Cronbach, Lee J. (1990). *Essentials of Psychological Testing*. Fifth edition. New York: Harper and Row Publishers.
- Gronlund, Norman E. (1973). *Preparing Criterion-Referenced Test for Classroom Instruction*. New York: The Macmillan Publishing Company.
- Lado, Robert. (1961). *Language Testing*. London: Longman Group Limited.
- Muhaimin, dkk. (1996). *Strategi Belajar Mengajar*. Surabaya: CV. Citra Media.

- Oemar Hamalik. (1995). *Kurikulum dan Pembelajaran*. Jakarta: Bumi Aksara.
- Parera, Jos D. (1997). *Lingustik Edukasional*. Jakarta: Erlangga.
- Subino. (1987). *Konstruksi Dan Analisi Tes Suatu Pengantar Kepada Teori Tes Dan Pengukuran*. Jakarta: Depdikbud.
- Suharsimi Arikunto. (1992). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Sutardi Wirasasmita. (1998). *Teknik Penyusunan dan Analisis Tes Prestasi Belajar dengan Pengembangan Tes Prestasi Belajar Bahasa Indonesia*. Bandung: IKIP.