

PENGEMBANGAN INSTRUMEN PENGUKURAN HASIL PEMBELAJARAN MATA KULIAH *PRONUNCIATION*

Jamilah dan Ari Purnawan
Jurusan Pendidikan Bahasa Inggris
FBS Universitas Negeri Yogyakarta
Email: aripurnawan@yahoo.co.id

ABSTRACT

The study aims at developing a valid and reliable test to measure students' English pronunciation mastery. To meet the objective, two kinds of instruments were developed, namely recognition and production tests. The instruments were tried out to students from three pronunciation classes, and the results were then analyzed through validity and reliability tests and an item analysis. Statistical analyses showed that the instruments were unsatisfactory and needed revisions. After some revisions and try-outs, the revised instruments proved to be valid and reliable with reliability coefficients of 0.940 for the production test and 0.889 for the recognition test, and difficulty indices of 0.538 (moderate) and 0.546 (moderate) respectively.

Keywords: pronunciation, validity, reliability, difficulty indices

A. PENDAHULUAN

Dalam matakuliah *Pronunciation* di Jurusan Pendidikan Bahasa Inggris FBS UNY, masing-masing dosen memberikan perangkat evaluasi yang berbeda-beda sesuai dengan keyakinan dan kemampuan mengukur dosen yang bersangkutan. Silabus, kisi-kisi bahan, dan bahan ajar serta aktivitas belajar-mengajar di kelas umumnya sama, namun seringkali terjadi perbedaan yang menyolok antara penilaian belajar satu kelas dengan kelas lain yang diajar oleh dosen lain. Bisa jadi ini merupakan hasil dari akumulasi dari berbagai unsur seperti kemampuan siswa, kemampuan dosen, atau aspek lain, namun dapat pula hal di atas disebabkan oleh alat ukur penilaian yang saling berbeda dan tidak baku, baik segi kelayakan, keahlian, maupun kepercayaan tesnya. Berkaitan dengan belum adanya alat ukur yang baku untuk satu mata kuliah yang diajar oleh beberapa dosen ini, maka penelitian pengembangan instrumen pengukuran ini perlu dilakukan, sehingga akan didapatkan alat ukur yang lebih objektif dan adil, tidak menguntungkan bagi satu pihak dan merugikan pihak lain.

Sebuah tes prestasi yang baku pada prinsipnya dapat mengukur suatu aspek dengan tepat. Subjek yang berbeda pada tempat yang berbeda dan pada waktu yang berbeda dapat dikenai tes tersebut dan hasilnya juga dapat mencerminkan tingkat pencapaian suatu konstruk yang dimiliki oleh seorang subjek. Untuk mendapatkan kebakuan semacam ini tidak mudah. Tes baku memerlukan proses yang panjang untuk mencapainya. Perangkat kebakuan tes, antara lain metode baku untuk penyusunannya, penyelenggaraanya, penyekorannya, dan pelaporan hasilnya. Dalam pembuatannya, tes tersebut harus melalui proses standarisasi, yang mencakup ujicoba, analisis untuk mendapatkan koefisien tertentu dengan teknik tertentu, penulisan kembali atau perbaikan, dan pengujian kembali. Kebakuan baru didapatkan setelah tes versi terbaik setelah direvisi itu diterapkan pada *norming sample*, yaitu sampel sejumlah sangat besar dari berbagai wilayah dengan kondisi dan tingkat serupa dengan subjek ujicoba, dan hasil yang didapatkan merupakan hasil untuk komparasi.

Langkah demikian merupakan langkah wajib yang harus dilakukan oleh lembaga

pengetesan komersial atau formal yang menerbitkan tes baku untuk area yang luas. Dalam lingkup penggunaan tes yang lebih sempit, apabila seorang pengajar ingin menghasilkan tes yang mempunyai tingkat validitas dan reliabilitas bagus dan hasilnya dapat mengukur pencapaian belajar untuk subjek yang ada dalam jangkauannya, prosedur penerapan *norming sample* antarwilayah tentu menyulitkan, karena wilayah penggunaan atau jangkauan kebakuan menentukan keluasan *norming sample* tersebut. Dalam penelitian ini, wilayah keluasan atas kebakuan instrumen adalah untuk seluruh kelas *pronunciation* di lingkup satu jurusan, sehingga wilayah uji coba dan *norming sample* tidak sampai ke luar jurusan atau ke jurusan Bahasa Inggris di universitas lain. Dengan demikian kebakuan yang dicapai oleh instrumen hasil penelitian ini hanya berlaku untuk mengukur hasil belajar mata kuliah *Pronunciation* di Jurusan Pendidikan Bahasa Inggris FBS UNY saja, dan tidak dapat digunakan untuk dengan tepat mengukur pembelajaran mata kuliah yang serupa yang dicapai oleh subjek dari perguruan tinggi lain, yang mungkin memakai silabus yang berbeda, menerapkan proses belajar yang berbeda, dan memakai sistem penilaian yang berbeda.

Baik atau tidaknya sebuah instrumen pengukuran pada dasarnya ditentukan oleh hasil pengamatan dan pengujian terhadap instrumen tersebut. Dua ukuran pokok yang selama ini dipakai sebagai penentu adalah validitas dan reliabilitas. Untuk mengetahui lebih jauh status tes tersebut, seringkali dilakukan serangkaian uji lainnya, yaitu tingkat kesulitan, daya beda, dan analisis distraktor.

Validitas merupakan kriteria penentu kebaikan suatu tes. Validitas dapat diartikan sebagai "... *the degree to which the test actually measures what is intended to measure and what it purports to measure*" (Brown, 1987:221; Cohen, 1994:38). Tes *pronunciation* yang valid hanya mengukur pengetahuan tentang dan keterampilan dalam mengidentifikasi bunyi, stress, intonasi, dan pengucapannya menurut standar tertentu. Skor

yang dicapai dalam tes *pronunciation* tidak ditentukan oleh aspek lain seperti pemahaman struktur kalimat, kemampuan *reading comprehension*, atau variabel lain yang relevansinya dengan *pronunciation* kurang jelas. Apabila aspek atau variabel lain masih mewarnai pengukuran suatu konstruk atau subjek, validitas hasil pengukuran itu masih rendah.

Prosedur validasi tes, baik yang berupa tes standar maupun tes yang digunakan untuk lingkup kelas, merupakan hal yang kompleks dan memerlukan pengetahuan dan keterampilan. Validasi teoritik yang paling banyak diketahui dan dipakai adalah *content validity* (validitas isi) dan *construct validity* (validitas konstruk), sedangkan validitas empirik dicapai dengan langkah korelasi.

Validitas isi ditentukan dengan melihat keterwakilan dan kesesuaian isi atau tujuan pembelajaran dengan butir-butir tes (Cohen, 1994:38). Langkah yang biasa dilakukan adalah dengan mencermati tes butir demi butir, mengecek tabel spesifikasi atau kisi-kisi, dan melakukan review atas isi pembelajaran sesuai dengan tujuan yang akan dicapai. Sebuah instrumen dapat dikatakan valid jika aspek-aspek, tujuan, dan isi pembelajaran tercakup atau terwakili dalam tes. Tes yang isinya tidak menyampel atau mengukur aspek isi atau tujuan pembelajaran dinyatakan sebagai tes dengan validitas isi yang rendah. Beberapa langkah yang perlu dilakukan untuk mendapatkan tes dengan validitas isi yang baik adalah (1) membuat daftar seluruh cakupan topik yang akan dicapai berdasarkan isi materi yang disampaikan, (2) pembobotan topik dalam aspek pembelajaran, (3) membangun tabel spesifikasi, dan (4) menyusun instrumen berdasar tabel tersebut. Semakin dekat bentuk instrumen dengan pembobotan unsur-unsur dalam tabel spesifikasi, semakin besar kemungkinan didapatkan respon yang mempunyai validitas yang tinggi. Dengan demikian, validitas suatu instrumen tercapai apabila aspek yang akan diukur memang ada dan terwakili secara proporsional.

Reliabilitas merupakan ukuran berikutnya. Sebuah tes yang reliabel adalah tes yang menghasilkan skor yang konsisten (Tinambunan, 1988). Konsisten maksudnya adalah apabila hasil yang dicapai seorang peserta apabila tingkat penguasaannya ajeg akan menunjukkan skor, peringkat, atau tingkat yang kurang lebih sama. Salah satu cara untuk mengetahui reliabilitas suatu instrumen adalah dengan metode konsistensi internal, yang mengindikasikan keajegan skor tes apabila diuji antarbagian tes tersebut. Maksudnya, seorang peserta yang menjawab suatu butir dengan benar diasumsikan cenderung menjawab butir-butir lainnya dengan benar pula. Dengan demikian, butir-butir tes tersebut saling dikorelasikan. Hasil korelasi yang tinggi mengindikasikan bahwa tes tersebut mempunyai konsistensi internal, dan dengan demikian reliabel sebagai suatu perangkat pengukuran. Cara-cara yang biasa digunakan untuk mengukur konsistensi internal suatu instrumen adalah teknik belah dua dan formula Kuder-Richardson (KR). Apabila suatu instrumen pengukur terdiri dari dua set/jenis soal untuk mengukur konstruk yang sama, dapat dilakukan pula uji keahihan faktor (Hadi, 1996), dengan satu jenis merupakan faktor ke satu dan jenis yang lain merupakan faktor ke dua. Respon peserta tes kedua faktor tersebut kemudian dikorelasikan. Korelasi yang tinggi mencerminkan adanya keajegan/reliabilitas instrumen.

Selain validitas dan reliabilitas, tingkat kesulitan dan uji daya beda juga dapat dilakukan untuk melihat efektivitas sebuah instrumen. Tingkat kesulitan menyangkut rasio antara penjawab suatu butir yang benar dengan jumlah seluruh jawaban butir tersebut. Semakin sedikit penjawab benar untuk sebuah butir tertentu, semakin rendah indeks kesulitannya. Sebuah butir yang mempunyai indeks ekstrim menjadi kurang bermakna, karena butir tersebut kurang dapat memisahkan antara peserta yang menguasai *subject matter* dan yang kurang menguasainya. Uji daya beda berkaitan dengan kemampuan butir untuk membedakan antara kelompok skor tinggi dengan kelompok skor

rendah. Bentangan indeks daya beda adalah -1 sampai +1. Daya beda +1 berarti semua anggota kelompok berskor tinggi menjawab benar butir tersebut, sedangkan semua anggota berskor rendah tidak ada yang menjawab benar untuk butir tersebut. Sebaiknya, daya beda -1 berarti semua anggota kelompok berskor tinggi menjawab salah butir tersebut, sedangkan semua anggota kelompok berskor rendah menjawab benar. Tentu saja butir soal dengan daya beda negatif ini tidak berguna, bahkan menyesatkan. Agar sebuah butir dapat disebut berdaya beda baik, minimal indeksnya adalah 0,5, sedangkan yang kurang dari 0,25, apalagi negatif, tidak merupakan butir yang baik (Hasan dan Zainul, 1991:132). Azwar (1996) memberikan batasan yang sedikit berbeda, yaitu 0,2 atau kurang baru merupakan butir tidak berguna. Untuk tujuan pembakuan instrument, tentu batas minimal ini bukanlah harga yang ideal. Bila masih memungkinkan, butir-butir yang ada disekitar titik kritis tersebut direvisi untuk mendapatkan butir yang lebih baik.

Sebagai rangkaian tak terpisahkan dalam proses belajar bahasa, *Pronunciation* diajarkan pada semester pertama di hamper semua jurusan Bahasa Inggris, termasuk di Jurusan Pendidikan Bahasa Inggris FBS UNY. *Pronunciation* termasuk salah satu mata kuliah yang mendasari kuliah-kuliah *skill* seperti *Speaking* dan *Listening*. Karena muara belajar *pronunciation* adalah pada kemampuan pengucapan bunyi bahasa yang benar menurut kaidah yang disepakati para penggunanya, pembelajaran keterampilan berbicara yang diwujudkan dalam mata kuliah *Speaking* sangat dipengaruhi oleh proses dan hasil pembelajaran *pronunciation*. Demikian juga dengan *Listening*. Dalam belajar *Listening*, konsep bunyi dan pemahaman makna sangat ditentukan oleh persepsi dan konsep abstrak tentang bunyi yang ada di otak pembelajar.

Di antara kedua unsur pokok tersebut, kemampuan mengucapkan bunyi bahasa dan kemampuan memahami konsep bunyi merupakan unsur pokok dalam belajar *pronunciation*. Karenanya, pengetesan

pronunciation juga hendaknya mencakup kedua unsur tersebut. Pengetesan atas salah satu aspek saja mungkin merupakan bentuk yang paling cocok untuk mengukur aspek yang sangat spesifik dari *pronunciation*, sedangkan kombinasi keduanya akan lebih mewakili aspek belajar *pronunciation* secara keseluruhan.

Aspek-aspek yang dipelajari dalam mata kuliah *Pronunciation* pada dasarnya adalah bunyi (*sounds*), tekanan (*stress*), dan intonasi (*intonation*). Pada bagian bunyi, mahasiswa belajar mengenali dan mengucapkan berbagai bunyi dalam bahasa Inggris. Pada bagian *stress*, mahasiswa belajar mengenali dan mengucapkan kata atau kelompok kata dengan penekanan pada suku kata tertentu. Penekanan pada suku kata tertentu ini tidak dikenal dalam bahasa ibu, sedangkan pola *stressing* pada kata atau frase bahasa Inggris sudah baku, bahkan *stress* ini merupakan salah satu unsur pembeda arti.

Demikian juga dengan intonasi kalimat, ditemui cara pemberian intonasi yang tidak sama dengan pola intonasi bahasa yang biasa dipakai mahasiswa sehari-hari di rumah. Intonasi pada kalimat tanya, misalnya, dapat berbeda-beda tergantung tujuan dan jenis kalimat tanyanya. Intonasi juga dapat mengakibatkan perubahan makna, misalnya kalimat berita dapat berubah menjadi kalimat tanya hanya dengan mengubah intonasi.

Selain beberapa aspek di atas beberapa fenomena penggabungan bunyi yang mengakibatkan terjadinya perubahan juga merupakan bahasan dalam *pronunciation*. *Assimilation*, *deletion*, *juncture*, dan *contraction* adalah perubahan bunyi yang perlu diketahui oleh pembelajar.

Bagi pembelajar yang mempunyai latar belakang bahasa ibu yang sangat berbeda dengan bahasa Inggris, belajar *pronunciation* bahasa Inggris merupakan masalah tersendiri. Setiap bahasa mempunyai kosa kata dan tata cara pengucapan sendiri-sendiri, dan terkadang bunyi-bunyi tertentu sangat sulit diucapkan karena seseorang pembelajar tidak terbiasa mengucapkannya. Orang Indonesia yang belajar *pronunciation* bahasa Inggris biasanya

mempunyai cukup banyak masalah karena adanya berbagai perbedaan seperti di atas.

Secara fonetis, bahasa Inggris sangat berbeda dengan bahasa-bahasa yang dipakai masyarakat Indonesia pada umumnya. Bahasa Inggris mempunyai 24 bunyi konsonan, 12 vokal dan 8 diftong (Homby, 1995:1429). Bunyi-bunyi konsonan, vokal dan diftong tersebut sebagian merupakan bunyi khas yang tidak ditemui pada bahasa ibu, seperti bunyi /æ/, bunyi-bunyi vokal panjang dan pendek seperti /ç:, I,eY/, beberapa bunyi konsonan seperti /ð, ʔ/ atau bunyi /+ʰ, dʰ, t+ʰ/. Bahasa-bahasa Indonesia, Jawa dan beberapa bahasa daerah yang lebih dulu dikuasai oleh pembelajar *pronunciation* tidak mengenal bunyi-bunyi tersebut. Oleh karena itu, dimungkinkan pembelajar akan mengalami kesulitan mengucapkan bunyi-bunyi tersebut, baik bunyi tunggal secara terpisah maupun bunyi dalam konteks. Dengan demikian, salah satu penentu keberhasilan pembelajaran *pronunciation* adalah kemampuan mengenali dan memproduksi bunyi-bunyi semacam itu, di samping berbagai aspek lainnya.

Pengukuran hasil belajar *pronunciation* tentunya juga menyangkut dua aspek utama di atas *recognition* dan *production*. Hal ini merupakan keharusan karena kedua aspek tersebut menentukan pemahaman dan kemampuan atau keterampilan memproduksi bunyi atau ujaran. Pengukuran dengan *recognition test* memberikan hasil sejauh mana seorang pembelajar mengetahui dan menerapkan *rules* atau tata aturan dasar pengucapan bunyi. Tes semacam ini dapat dilakukan dengan bantuan media audio, di mana pembelajar mendengarkan bunyi dan kemudian mengidentifikasi bunyi yang didengarnya tersebut. Karena *pronunciation* berkaitan dengan kata, sedangkan kata dibedakan dari bunyi –untuk lisan– atau dari ejaan –untuk tulisan–, tes *recognition* dapat juga dilakukan dengan bentuk tertulis, di mana pembelajar mengidentifikasi ejaan kata untuk menentukan bentuk pengucapan yang benar.

Kemudian pengukuran dengan *production test* sangat penting karena

kemampuan menandai atau *recognition* tidak mencerminkan kemampuan mengucapkannya. Sebagai contoh, seorang yang bisu mungkin saja bisa mengerjakan tes *recognition* untuk *pronunciation* dengan sempurna. Tingginya skor itu tentu saja tidak dapat digunakan untuk menyimpulkan bahwa orang tersebut mempunyai *pronunciation* yang bagus. Karena itulah diperlukan jenis tes lain yang mengukur sejauh mana seseorang dapat mengucapkan bunyi, kata, frase dan kalimat dengan tata bunyi yang benar.

B. METODE PENGEMBANGAN INSTRUMEN

Prosedur yang dilakukan untuk mengembangkan instrumen baku pengukuran hasil belajar *pronunciation* yang dapat digunakan di lingkup jurusan Pendidikan Bahasa Inggris FBS UNY meliputi : (1) Perencanaan dan penyusunan instrumen, (2) Uji coba instrumen, (3) Analisis hasil pengujian uji coba pertama, (4) Revisi instrumen, dan (5) Analisis pengujian hasil uji coba kedua.

1) Perencanaan dan Penyusunan Instrumen

Langkah-langkah yang dilakukan untuk menyusun instrumen pengukuran hasil belajar *pronunciation* adalah mengembangkan kompetensi dan indikator hasil pembelajaran yang diturunkan dari silabus dan referensi lain yang berhubungan dengan pembelajaran *pronunciation*. Dari indikator-indikator tersebut kemudian dibuat tabel yang sesuai. Langkah berikutnya adalah membuat butir-butir instrumennya lengkap dengan kriteria pengukuran dan penelaahan butir-butir instrumen.

Aspek-aspek yang dikembangkan dalam butir-butir tes mencakup pengetahuan dan keterampilan dalam mengenali dan mengucapkan (*recognition and production*). Pengetahuan dan keterampilan ini berkaitan dengan aspek-aspek bunyi-bunyi vokal, konsonan dan diftong dalam bahasa Inggris, memberi tekanan (*stress*) pada kata dan

kelompok kata (*frase*), pengucapan bunyi lemah dan kuat, intonasi, dan perubahan-perubahan bunyi yang disebabkan proses fonologis dan bentuk-bentuk atau ujaran-ujaran khusus dalam konteks yang berbeda-beda.

Tes yang dikembangkan ada dua buah, yaitu satu set tes untuk mengukur kemampuan mengenali (*recognition test*) dalam bentuk tes tertulis, dan satu set tes untuk mengukur kemampuan mengucapkan (*production test*) dalam bentuk tes lisan. Masing-masing peserta tes mengerjakan kedua macam tes tersebut. Kedua bentuk tes ini diberikan kepada mahasiswa secara berturut-turut; segera setelah mengerjakan tes tertulis kemudian mahasiswa mengerjakan tes lisan.

Tes kemampuan mengenali dibuat dalam bentuk pilihan ganda yang terdiri dari 40 butir soal, masing-masing dengan 4 pilihan dengan 1 jawaban benar dan 3 distraktor. Mahasiswa mengerjakan tes dengan cara memberi tanda pada jawaban yang menurut mereka paling tepat. Agar hasil tes dapat terjaga dari bias yang disebabkan oleh kecurangan peserta, susunan ketika tes diberikan dibuat sedemikian rupa sehingga memperkecil kesempatan untuk menyontek dan bekerjasama.

Tes kemampuan mengucapkan dilakukan dengan bantuan *tape recorder*. Setiap peserta diberi lembar tes yang berisi kata, frase, dan kalimat yang harus mereka ucapkan dan ucapan mereka direkam dengan pita kaset. Tabulasi dari hasil rekaman dilakukan dengan cara melakukan pemutaran hasil rekaman. Hasilnya dituliskan dalam lembaran kertas yang berisi pilihan benar/salah untuk masing-masing ekspresi.

Berikut adalah kisi-kisi instrumen pengukuran yang meliputi *recognition test* dan *production test*. Kisi-kisi instrumen untuk kedua tes tersebut diturunkan dari teori tentang aspek-aspek belajar *pronunciation* dan silabus perkuliahan *pronunciation* yang dipakai sebagai acuan bahan pembelajaran selama kuliah berlangsung. Kisi-kisi tersebut kemudian disusun dalam indikator dan subindikator seperti dalam Tabel 1 dan Tabel 2.

Tabel 1. Kisi-kisi Instrumen *Recognition Test*

No.	Indikator	Subindikator	Jumlah Butir	Nomor Butir
1.	<i>Sounds</i>	<i>a. Vowels</i>	12	2,4,9,11,13,16,22,25,27,33, 38, 40
		<i>b. Consonants</i>	11	1,3,7,10,20,21,23,28,32,35, 36
		<i>c. Diphthongs</i>	6	5,14,18,31,34,37
2.	<i>Stress</i>	<i>a. Words stress</i>	3	8, 12, 19
		<i>b. Sentence stress</i>	2	17 6,
		<i>c. Weak and Strong forms</i>	1	26
3.	<i>Intonation</i>	<i>a. Statements</i>	1	30
		<i>b. Questions</i>	1	24
4.	<i>Other aspects of connected speech</i>	<i>a. Assilimilation</i>	1	15
		<i>b. Ellision</i>	1	29
		<i>c. Intrusion</i>	1	39
	Total		40	40

Instrumen kemudian disusun dengan berpedoman pada kisi-kisi di atas. Format dan tata cara penulisan butir, pemilihan kalimat yang baik, pemilihan distraktor, pengacakan butir, dan teknik pembuatan butir-butir soal pilihan ganda dilakukan dengan merujuk pada tata cara penulisan butir yang dikemukakan

oleh para pakar penyusunan tes, seperti Cohen (1994:62); Lado (1977:46); dan Stanly (1972:232). Instrumen yang telah disusun kemudian diseminarkan untuk mendapatkan masukan dan direvisi sesuai dengan masukan dan saran untuk perbaikan.

Tabel 2. Kisi-kisi Instrumen *Production Test*

No.	Indikator	Subindikator	Jumlah Butir	Nomor Butir
1.	<i>Sounds</i>	<i>a. Vowels</i>	13	7,8,13,14,31,37,41,54,58,64, 71,73,75
		<i>b. Consonants</i>	7	6,11,26,33,45,60,61
		<i>c. Diphthongs</i>	16	1,2,12,27,28,32,42,49,51,53, 55,57,66,67,74,76
2.	<i>Stress</i>	<i>a. Words stress</i>	14	3,4,5,10,15,19,20,21,24,25,29,69, 70,72
		<i>b. Sentence stress</i>	5	30,38,43,52,68
		<i>c. Weak and Strong forms</i>	5	44,47,48,50,80
3.	<i>Intonation</i>	<i>a. Statements</i>	6	16,22,23,24,36,40
		<i>b. Questions</i>	4	17,18,35,39
4.	<i>Other aspects of connected speech</i>	<i>a. Assilimilation</i>	4	9,62,63,65
		<i>b. Ellision</i>	3	56,77,78
		<i>c. Intrusion</i>	3	46,59,79
	Total		80	80

2) Uji Coba Instrumen

Instrumen yang telah tersusun dan telah diseminarkan kemudian diujicobakan pada sampel yang telah ditentukan. Uji coba dilakukan dalam 2 tahap. Tahap kedua dilakukan setelah dilakukan revisi terhadap butir-butir instrumen yang gugur dalam uji coba pertama. Revisi dilakukan dengan cara melakukan perubahan *stem* soal dan *options* untuk butir-butir tersebut, sedangkan aspek

atau konstruk pada subindikator yang diukur oleh butir soal tetap sama dengan aspek pada soal semula.

Uji coba pertama dilakukan pada bulan Nopember 2004, yaitu setelah para mahasiswa belajar *pronunciation* selama 11 atau 12 kali tatap muka (90-100% cakupan bahan). Skor yang dicapai dalam uji coba instrumen disajikan pada Tabel 3 berikut.

Tabel 3. Skor Ujicoba Tes Tahap Pertama

Hasil	Uji Pertama		Uji Kedua	
	Recognition	Production	Recognition	Production
Jumlah peserta	54	54	54	54
Skor tertinggi	40	78	40	79
Skor terendah	9	8	9	7
Rerata	21,59	43,28	21,00	43,22

3) Analisis Hasil Pengetesan Uji Coba Pertama

a). Uji Validitas Tes

Uji validitas meliputi dua hal, yaitu validitas isi dan validitas teoritik. Validitas isi dimintakan pada seorang dosen jurusan Pendidikan Bahasa Inggris FBS UNY, yang membandingkan antara silabus, kisi-kisi, dan instrumennya. Hasil kesimpulan yang diperoleh menunjukkan bahwa kisi-kisi tes

sudah mencakup semua aspek yang dituangkan dalam silabus perkuliahan, dan butir-butir instrumen yang disusun sudah sesuai dengan kisi-kisinya. Sebagai bahan pembandingan dilakukan *polling* terhadap peserta tes mengenai kesesuaian tes sebagai alat ukur untuk mengukur hasil belajar *pronunciation*. Tabel 4 di bawah ini menunjukkan hasil pendapat peserta.

Tabel 4. Pendapat Peserta tentang Kesesuaian Tes Tahap Pertama

No	Pendapat	Jumlah	Persen
1	Sesuai	32	59,26
2	Cukup sesuai	12	22,22
3	Kurang sesuai	6	11,11
	Total	54	100

Untuk mendukung data ini, dilakukan uji validitas teoritik. Uji validitas menghasilkan

luaran analisis sebagai tabel berikut.

Tabel 5. Hasil Uji Validitas Tahap Pertama

No	Tes	Gugur		Total Butir
		Jumlah	Nomor	
1.	Recognition	8	8, 11, 16, 20, 23, 30, 36, 39	40
2.	Production	11	5, 10, 15, 21, 24, 29, 37, 57, 63, 67, 72	80

b). Uji Reabilitas Tes

Uji reabilitas tes untuk masing-masing instrument dilakukan dengan teknik yang

berbeda, yaitu KR-20 dan belah dua. Hasil uji reliabilitas disajikan pada Tabel 6.

Tabel 6. Hasil Uji Reliabilitas Tahap Pertama

Tes	KR-20		Belah Dua		Status
	Koefisien	p.	Koefisien	p.	
Recognition	0,873	0,000	0,798	0,000	Reliabel
Production	0,925	0,000	0,764	0,000	Reliabel

c). Analisis Butir

Analisis butir meliputi uji daya beda dan tingkat kesulitan butir. Uji daya beda diterapkan untuk masing-masing jenis tes. Uji daya beda untuk tes *recognition* menghasilkan 32 butir yang mempunyai kemampuan membedakan peserta tes yang masuk kelompok pandai dan kurang pandai, sedangkan 8 butir memerlukan revisi. Daya beda tertinggi adalah 0,612 dan daya beda terendah (untuk butir tidak gugur) adalah 0,308.

Uji daya beda untuk tes *production* menghasilkan 71 butir tes yang mempunyai kemampuan membedakan peserta tes yang masuk kelompok pandai dan kurang pandai, sedangkan 9 butir mempunyai koefisien daya beda yang kurang dari 0,250, yaitu butir nomor 5, 10, 15, 24, 29, 37, 57, 63, dan 72, sehingga memerlukan revisi. Daya beda tertinggi adalah 0,567 dan daya beda terendah (untuk butir tidak gugur) adalah 0,266.

Uji tingkat kesulitan butir dan tingkat kesulitan naskah soal untuk tiap jenis tes adalah sebagai berikut : Untuk tes *recognition* menunjukkan bahwa dari 40 butir soal, tingkat kesulitan butir soal bervariasi dari 0,389 sampai 0,667 dengan rerata 0,5542. Untuk tes *production*, hasil analisis menunjukkan bahwa

dari 80 butir, tingkat kesulitan butir bervariasi dari 0,407 sampai 0,704, dengan rerata 0,5409.

4) Revisi Instrumen

Uji validitas menunjukkan ada 8 butir tes yang gugur untuk tes *recognition*, dan 11 butir gugur untuk tes *production*. Sedangkan uji reliabilitas untuk butir sah menunjukkan bahwa semua butir tersebut reliabel. Hasil analisis daya beda menunjukkan adanya beberapa butir yang gugur dan butir-butir yang gugur adalah butir yang gugur pada uji validitas. Butir-butir yang gugur tersebut kemudian direvisi dengan beberapa cara.

Untuk tes *recognition* bentuk pilihan ganda, revisi dilakukan dengan cara merubah kalimat pada *stem*. Kebetulan semua butir yang gugur tersebut memakai bentuk *stem* negatif atau perkecualian, yaitu diakhiri dengan kata *except* (kecuali). Hal ini diduga merupakan salah satu penyebab gugurnya butir, meskipun tidak semua *stem* yang menggunakan kata *except* gugur. Ada butir lain yang menggunakan kata perkecualian tapi ternyata tidak gugur. Karena itu, selain mengubah kata *except*, perbaikan juga dilakukan dengan mengubah pilihan jawaban, yang diduga juga merupakan penyebab gugurnya butir. Pengubahan pilihan

jawaban dilakukan dengan cara mencari kata yang lebih dikenal mahasiswa. Untuk butir tes *production*, revisi dilakukan dengan cara mengganti kata, frase, atau kalimat yang harus diucapkan mahasiswa dengan kata, frase, atau kalimat lain yang mungkin lebih dikenal atau lebih umum dan sering digunakan. Hal ini mengurangi aspek upaya *guessing* ucapan, yang menghasilkan pengukuran yang kurang tepat.

Tingkat kesulitan naskah soal *recognition* adalah 0,5398 dan tingkat kesulitan naskah soal *production* adalah 0,5409. Harga di atas menunjukkan butir tes mempunyai kesulitan yang sedang. Namun bila hal ini dijadikan ukuran kelulusan, diduga akan banyak peserta yang tidak lulus atau mempunyai nilai rendah. Analisis ulang terhadap butir hasil revisi menjadi menarik untuk dilakukan, dan hasil kedua versi tes ini dapat dibandingkan.

5) Analisis Hasil Pengetesan Hasil Uji Coba Kedua

a) Uji Validitas Tes

Seperti pada uji pertama, uji validitas meliputi dua hal, yaitu validitas isi dan validitas teoritik. Validitas isi dimintakan pada seorang dosen Jurusan Pendidikan Bahasa Inggris FBS UNY yang juga menjadi validator untuk tes tahap pertama. Untuk validasi ini dilakukan dengan membandingkan antara silabus, kisi-kisi, dan instrumennya. Hasil kesimpulan yang diperoleh menunjukkan bahwa kisi-kisi tes sudah mencakup semua aspek yang dituangkan dalam silabus perkuliahan, dan butir-butir instrumen yang disusun sudah sesuai dengan kisi-kisinya. *Polling* kembali dilakukan terhadap peserta tes untuk meminta pendapat mereka mengenai kesesuaian tes. Tabel berikut menyajikan hasil *polling* tahap kedua.

Tabel 7. Pendapat Peserta tentang Kesesuaian Tes Tahap Kedua

No.	Pendapat	Jumlah	Persen
1.	Sesuai	28	51,85
2.	Cukup sesuai	19	35,19
3.	Kurang sesuai	4	7,41
	Total	54	100

Untuk mendukung data ini dilakukan uji validitas teoritik kembali. Uji validitas teoritik memberi hasil seperti dalam Tabel 8.

Tabel 8. Hasil Uji Validitas Tahap Kedua

No	Tes	Gugur		Total Butir
		Jumlah	Nomor	
1.	Recognition	4	7, 20, 23, 36	40
2.	Production	5	10, 37, 57, 63, 72	80

Keempat butir yang gugur dalam tes *recognition* ini adalah butir-butir yang gugur pada uji coba pertama, sedangkan 4 butir lain yang gugur pada uji pertama sekarang dinyatakan valid. Pada tes *production*, butir-butir yang tetap gugur setelah direvisi ini adalah juga butir yang telah gugur pada ujicoba pertama, sedang

dari 11 butir gugur pada ujicoba pertama tersebut, 6 butir menjadi valid setelah direvisi.

b) Uji Reliabilitas Tes

Uji reliabilitas tes untuk masing-masing instrumen pada tahap kedua ini

dilakukan dengan 3 teknik yang berbeda, yaitu KR 20, belah dua, dan kesahihan faktor. Rangkuman hasil analisis disajikan pada Tabel 9.

Tabel 9. Hasil Uji Reliabilitas Tahap Kedua

Tes	KR-20		Belah Dua		Kesahihan Faktor	Status
	Koefisien	p.	Koefisien	p.		
<i>Recognition</i>	0,889	0,000	0,800	0,000	0,818	Reliabel
<i>Production</i>	0,940	0,000	0,831	0,000	0,818	Reliabel

c) Analisis Butir

Seperti pada ujicoba pertama, analisis butir meliputi uji daya beda dan tingkat kesulitan butir. Uji daya beda diterapkan untuk masing-masing jenis tes. Uji daya beda untuk tes *recognition* menghasilkan 36 butir yang mempunyai kemampuan membedakan peserta tes yang masuk kelompok pandai dan kurang pandai, sedangkan 4 butir mempunyai koefisien daya beda yang kurang dari 0,250, yaitu butir nomor 7, 20, 23, dan 36. Daya beda tertinggi adalah 0,662 dan daya beda terendah (untuk butir tidak gugur) adalah 0,316.

Uji daya beda untuk tes *production* menghasilkan 75 butir tes yang mempunyai kemampuan membedakan peserta tes yang masuk kelompok pandai dan kurang pandai, sedangkan 5 butir mempunyai koefisien daya beda yang kurang dari 0,250, yaitu butir nomor 10, 37, 57, 63, dan 72. Daya beda tertinggi adalah 0,583 dan daya beda terendah (untuk butir tidak gugur) adalah 0,289.

Uji tingkat kesulitan butir dan tingkat kesulitan naskah soal untuk tiap jenis tes adalah sebagai berikut. Tes *recognition* menunjukkan bahwa dari 40 butir soal, tingkat kesulitan butir soal bervariasi dari 0,370 sampai 0,685, dengan rerata 0,5462. Untuk tes *production*, hasil analisis menunjukkan bahwa dari 80 butir, tingkat kesulitan butir bervariasi dari 0,4074 sampai 0,7037 dengan rerata 0,5377.

C. DISKUSI HASIL ANALISIS

Penyusunan kedua jenis tes di atas –*recognition* dan *production*– dilakukan dengan berdasar kisi-kisi yang disusun dari deskripsi teori tentang belajar *pronunciation*.

Butir-butir dikembangkan untuk melihat apakah mahasiswa telah menguasai teknik yang benar dalam pengucapan bunyi bahasa Inggris. Hasil pengujian terhadap instrumen pengukuran dalam penelitian ini menunjukkan bahwa instrumen tersebut dapat dijadikan alat ukur yang representatif dalam mengukur hasil pembelajaran yang dilakukan.

Dari beberapa butir yang terbukti gugur, dua alasan utama penyebab keguguran tersebut adalah soal terlalu sulit atau sebaliknya, terlalu mudah. Soal terlalu sulit dan terlalu mudah menjadi gugur karena tidak dapat berfungsi sebagaimana yang diharapkan, terutama dalam membedakan peserta tes yang pandai dan kurang pandai.

Satu temuan yang dapat dijadikan analisis lebih lanjut adalah bahwa semua butir tes *recognition* yang gugur memasukkan kata *except* (kecuali) dalam soal. Pemakaian kata *except* tersebut dapat jadi membuat salah tangkap peserta akan maksud dan pertanyaan tersebut. Pemakaian kata-kata seperti itu dalam sebuah naskah soal, apalagi dengan kemunculannya yang acak sering mengakibatkan peserta tes mengalami kebingungan. Peserta mungkin saja sebenarnya tahu jawaban yang benar untuk suatu butir, tetapi karena yang diinginkan adalah jawaban sebaliknya, peserta dapat tertipu memilih jawaban yang salah. Kelengahan seperti ini tentu saja bukan menjadi salah satu aspek pengukuran yang diinginkan.

Memang salah satu aspek penting dalam pengukuran *pronunciation* adalah ketepatan membaca ejaan, misalnya kata *through* dan *though*, meskipun sepintas mirip, cara mengucapkannya sangat berbeda.

Kelengahan dan kesalahan membaca atau mengidentifikasi kata di atas akan dapat membuat kesalahan *pronunciation*, dan pengukuran dengan cara seperti ini wajar atau bahkan merupakan salah satu aspek penting dalam pengukuran hasil belajar *pronunciation*.

Karenanya, dalam revisi butir-butir yang gugur dalam instrumen, jenis pertanyaan diubah dengan menghilangkan kata *except*, di samping kemudian diperlukan penyesuaian jawaban benar dan distraktornya. Pengubahan ini membuat instrumen gugur berkurang jumlahnya, meskipun kemudian tidak dapat dijadikan dasar penyimpulan bahwa penyebab gugurnya butir adalah adanya kata *except*. Hal ini terbukti dari banyaknya butir lain yang juga memakai kata *except* namun tidak gugur, dan sebaliknya, meskipun kata *except* sudah dihilangkan tetap saja ada butir yang gugur.

Satu alasan logis yang dapat menjelaskan fenomena di atas adalah bahwa memang aspek yang diukur oleh butir-butir gugur tersebut mengukur aspek yang sulit bagi kebanyakan peserta tes, sehingga hampir semua peserta tes salah menjawabnya, atau sebaliknya, aspek yang diukur adalah sesuatu yang sangat mudah bagi peserta, sehingga hampir semuanya menjawab dengan benar. Di satu sisi, gugurnya butir mengindikasikan tidak dapat dipakainya butir dalam pengukuran. Masalah muncul kemudian apabila aspek-aspek yang diukur dalam butir gugur tersebut adalah aspek pokok dan tidak dapat dihilangkan dari kesatuan naskah pengukuran. Atas pertimbangan itu, butir-butir pokok yang menang sangat penting dan merupakan cerminan indikator pokok dalam pelajaran *pronunciation* tetap dipertahankan, sedangkan butir gugur yang sudah terwakili oleh butir lain yang sah dapat dikeluarkan atau dihilangkan dari naskah soal.

Dalam kasus penyebab gugurnya beberapa soal dalam tes *production*, kejadian yang sama terjadi juga, yaitu mayoritas benar atau mayoritas salah mengucapkan kata, frase, atau kalimat dalam soal. Satu temuan yang juga menarik untuk dikaji lebih lanjut adalah bahwa kesalahan pengucapan kata-kata yang sudah

diserap dalam bahasa Indonesia mendominasi butir gugur tersebut. Sebagaimana diketahui, Bahasa Indonesia dalam perkembangannya banyak menyerap kata atau ungkapan dalam bahasa asing, terutama bahasa Inggris, yang diindonesiakan dengan beberapa perubahan atau penyesuaian fonetis.

Penyesuaian ini biasanya mengakibatkan pengucapan dan juga penulisan kata-kata serapan tersebut berubah atau berbeda dengan bentuk aslinya. Pada umumnya, pengubahan dilakukan dengan beberapa cara, misalnya dengan membaca huruf dalam kata asal seperti lazimnya membaca kata asli bahasa Indonesia, sehingga misalnya, kata *temperature* akan menjadi temperatur dan dibaca /tempYratur/, kata *special* menjadi spesial dan dibaca /spesial/, kata *computer* menjadi komputer dan dibaca /komputYr/. Biasanya *stressing* menjadi tidak penting lagi, dan pola *stressingnya* mengikuti pola yang biasa digunakan dalam logat bahasa Indonesia. Cara pengindonesiaan lainnya adalah dengan mengikuti cara pelafalan aslinya, sedangkan ejaannya mengikuti ucapan tersebut, seperti kata *design* akan ditulis disain karena mencoba mengikuti ucapan aslinya, meskipun kemudian juga ada perubahan dari pengucapan aslinya menjadi /disain/ atau /desain/.

Hal di atas ternyata sangat memengaruhi praktik pengucapan yang dilakukan para peserta tes pada instrumen tes *production*. Peserta tes cenderung salah mengucapkan kata-kata yang sudah diserap dalam bahasa Indonesia. Mereka cenderung mengikuti cara pengucapan kata yang bergaya Indonesia seperti bentuk serapannya. Apabila dikaitkan dengan kenyataan bahwa akhir-akhir ini jumlah kata serapan seperti ini meningkat dengan pesat dan frekuensi kemunculannya begitu tinggi karena mungkin kebetulan kata-kata serapan itu adalah kata-kata yang banyak digunakan di dunia ilmiah, golongan intelektual, dan pers, pengajar dan pembelajar *pronunciation* bahasa Inggris tampaknya perlu memperhatikan aspek ini sebagai salah satu kompetensi penting: tetap mampu melafalkan

kata dengan benar meskipun sudah ada versi Indonesia yang lebih terkenal.

Sementara itu revisi yang dilakukan atas kedua jenis instrumen pengukuran tersebut secara statistik telah meningkatkan koefisien validitas dan reliabilitas naskah soal. Peningkatan ini walaupun tidak drastis dan besar, telah membuktikan bahwa peningkatan mutu dan kesesuaian instrumen senantiasa dapat diupayakan sampai pada harga-harga tertentu yang diinginkan. Tingkat kesulitan naskah soal yang ada di kisaran 0,5 menunjukkan bahwa instrumen ini dapat dijadikan pengukur yang cukup baik dalam arti dapat membedakan peserta yang tidak menguasai bahan dan yang siap. Karena instrumen disusun berdasarkan kisi-kisi yang juga menjadi topik-topik bahasan dalam kuliah *pronunciation* di Jurusan Pendidikan Bahasa Inggris di FBS UNY, tentunya instrumen ini dapat tingkat ketuntasan bahan secara keseluruhan.

D. PENUTUP

Penelitian pengembangan instrumen pengukuran hasil belajar seperti ini merupakan penelitian yang tidak ada akhirnya, selalu ada keinginan untuk meningkatkan dan mengembangkan instrumen menuju status yang benar-benar ideal. Karenanya, untuk mendapatkan instrumen yang lebih baik lagi, peningkatan-peningkatan dapat dan perlu selalu dilakukan terhadap instrumen ini. Masing-masing butir yang dirasa masih ada kekurangan berdasar data empirik dalam penelitian ini dapat direvisi lagi kemudian diujicobakan lagi. Pengujian untuk kelompok dari perguruan tinggi lain mungkin diperlukan untuk mengetahui lebih lanjut apakah instrumen ini juga valid dan reliabel di sana.

Meskipun demikian, karena segala angka hasil analisis sangat terkait dengan kelompok subjek yang menjadi responden

ujicoba, pihak-pihak yang berkehendak menggunakan perangkat pengukuran ini untuk kelompok subjek yang berbeda dan mempunyai karakteristik yang tidak serupa disarankan berhati-hati. Dapat jadi instrumen ini tidak dapat memberikan hasil pengukuran yang diharapkan, hanya karena tingkat pengetahuan, keterampilan dan kemajuan peserta tes yang berbeda. Karenanya, penyesuaian mungkin perlu dilakukan terutama apabila kerangka acuan pengajaran *pronunciation* yang dipakai berbeda dengan yang dipakai oleh kelompok subjek penelitian ini.

E. DAFTAR PUSTAKA

- Azwar, S. 1996. *Tes Prestas: Fungsi dan Pengembangan Pengukuran Prestasi Belajar (Ed.2)*. Yogyakarta: Pustaka Pelajar.
- Brown, H.D. 1997. *Principles of Language Learning and Teaching (2nd ed)*. Englewood Cliffs: Prentice Hall.
- Brown, H.D. 2001. *Teaching by Principles: An interactive approach to language pedagogy*. New York: San Francisco State University.
- Burhan, N. 1995. *Penilaian dalam Pengajaran Bahasa dan Sastra (ed. 2)*. Yogyakarta: BPFE.
- Cohen, A.D. 1994. *Assessing Language Ability in the Classroom (2nd ed)*. Boston: Heinle & Heinle Publishers.
- Hadi S., Yuni P. 2001. *Seri Program Statistik (Manual dan Software Analisis Data Statistik)*. Yogyakarta: Bulaksumur.
- Hasan, S.H. dan A. Zainul. 1993. *Evaluasi Belajar*. Jakarta: Depdiknas.
- Hornby, A.S. 1995. *Oxford Advanced Learner's Dictionary of Current English (5th ed)*. Oxford: Oxford University Press.
- Kelly, G. 2000. *How to Teach Pronunciation*. Essex: Longman.