

BAB III

PERTIMBANGAN DALAM MENDESAIN TES BAHASA

3.1 Konsep Validitas

3.1.1 Membentuk Validitas

Konsep validitas (apakah mengerjakan test merupakan pengukuran?) dapatkah didekati dari sejumlah perspektif; hubungan antara yang ditafsirkan dalam sejumlah cara dalam literatur.

Penafsiran yang paling sangat menolong membentuk konsep validitas yang lebih tinggi. Anastasi (1982, p.153) berpendapat bahwa: isi, terkait dengan ukuran dan bentuk tidak sesuai dengan beda atau kategori yang terkoordinir secara logika. Sebaliknya, bentuk validitas adalah suatu konsep menyeluruh meliputi jenis lain.

Cronbach (1971, p.463) berkomentar bahwa: instrumen apa yang benar-benar mengukur?

Tingkatan tes yang dianggap lazim sebagai ukuran adalah suatu bentuk teoritis...masing-masing bentuk dikembangkan untuk menjelaskan dan mengorganisir konsistensi tanggapan diamati dan interlationship antar ukuran tingkah laku... pemusatan pada suatu yang lebih luas, lebih kronis dan lebih abstrak dari uraian bentuk tingkah laku...pengesahan memerlukan akumulasi informasi yang berangsur-angsur dari berbagai sumber. Data alami mempengaruhi penjelmaan dan pengembangan validitas.

Dia berargumentasi bahwa bentuk teoritis, ciri atau perilaku dapat digambarkan dalam menetapkan validitas untuk menekankan bentuk yang diukur oleh test tertentu (1982, p.155): test dipandang dari sudut data yang dikumpulkan... hanya merupakan penyelidikan yang empiris hubungan score test lain data eksternal.

Pandangan berbeda tentang data empiris eksternal dilihat untuk menetapkan ketercukupan suatu test untuk tujuan dimana (itu) dimaksudkan. Ada suatu kerangka teoritis untuk konstruksi test komunikatif, untuk menetapkan berdasar purbasangka betul validitas suatu test komunikatif, teori, dan untuk

menentukan bagaimana sukses test menjadi untuk ukuran. Perhatian terhadap aspek bentuk validitas yang tidak statistik mendekati ke paradigma yang teoritis test bahasa umum menguji unsur-unsur target bahasa secara terpisah dan berdasar purbasangka deliberation pada teori pengalaman. Validitas bersifat prediksi lebih konsisten dengan prinsip operationalism dan keinginan untuk sasaran ukuran eksternal.

Bentuk validitas dipandang dari perspektif statistik tidak terikat pada bentuk lain. Perhatian jauh lebih merupakan suatu hubungan antar suatu test dan kemampuan yang psikologis.

Untuk menetapkan validitas suatu test secara statistik, diperlukan perilaku kekuatan secara teoritis. Suatu prosedur yang diraikan oleh Campbell dan Fiske (1959) dan kemudiannya digunakan oleh Bachman dan Palmer (1981B). mereka menemukan poin-poin dengan memanfaatkan suatu acuan multi-method sebagai paradigma riset dalam membangun studi pengesahan. Mereka menemukan analisa faktoraplikasi konfirmasi untuk mengukur efek ciri dan metoda pada pengukuran kecakapan dan menyajikan suatu gambaran lebih jelas.

Percobaan perancangan acuan multi-method bersifat kritis mengarahkan test kecakapan bahasa adalah penyelidikan empiris yang bermanfaat untuk menjelaskan apa yang sudah diukur. Kesukaran dalam memanfaatkan teknik ini adalah berbeda kesalahan.

Berlawanan dengan pendapat bahwa ada suatu kebutuhan penting untuk membentuk purbasangka langkah disain test dan implementasi.

Cronbach (1971. Po 443) percaya bahwa: 'konstruksi test dimulai dari suatu teori tentang perilaku atau organisasi mental'. Davies (1977, p.63): bentuk test bersifat prediksi. Kelly (1978, p.8), berkomentar: pengembangan test yang sistematis memerlukan beberapa teori, informal, inexplicit, untuk memandu pemilihan awal isi item dan menarik perhatian.

Kita bisa menguraikan bentuk teoritis untuk mengukur, berdasarkan purbasangka langkah, data statistik yang tidak menghasilkan label konseptual. Kita berkewajiban menyelidiki bagaimana cukup suatu test sedang bekerja, melalui prosedur statistic yang tersedia.

3.1.2 Validitas Isi

Sebab kita kekurangan suatu teori bahasa yang cukup untuk digunakan, sehingga berdasar purbasangka mencoba untuk menentukan bentuk validitas test kecakapan yang menghubungkan dengan jelas kevaliditas isi. Dimensi test pencapaian semakin tampak dan sesuai dengan apa yang diketahui pencapaian itu, semakin mungkin untuk mempunyai isi dan bentuk validitas. Kita sering memperbincangkan tentang bentuk terminologi deskriptif komunikatif yang melibatkan keterkaitan isi dan pemenuhan isi. Kelly (1978, p.8) membenarkan bahwa hampir sepenuhnya konsep dengan bentuk validitas. Moller (1982b, p.68): perbedaan pad bentuk dan validitas isi dalam tes bahasa tidaklah selalu sangat ditandai, terutama sekali untuk test kecakapan bahasa umum.

Pembatasan waktu dan sumber daya untuk melibatkan konstruksi test yang digunakan di kelas adalah untuk memusatkan purbasangka pengesahan tugas test, perhatian tertentu harus difokuskan padda validitas isi untuk memastikan bahwa contoh aktifitas tercakup disuatu test.

Tujuan utama test komunikatif adalah untuk menyediakan suatu profil kecakapan siswa, menandakan terminologi defisiensi. Validitas isi dipertimbangkan untuk menuju keberhasilan tujuan karena itu adanya keterkaitan dengan tingkat pemilihan tugas test (bachman dan Palmer, 1981A).

Anastasi (1982, p.131) validitas isi digambarkan sebagai test yang sistematis untuk menentukan apakah contoh yang mewakili daerah perilaku terukur petunjuk bermanfaat untuk menetapkan validitas isi:

1. perilaku diuji secara sistematis dianalisa utuk memastikan bahwa semua aspek utama dicakup oleh materi test, dan dalam proposi yang benar.
2. pembahasan harus secara penuh diuraikan di depan, bukannya digambarkan setelah test disiapkan.
3. Validitas isi tergantung pada keterkaitan test individu menjawab area perilaku dalam pembahasan, bukannya pada keterkaitan isi item yang nyata.

J.B. Carroll (1961) melukiskan bahwa area bahasa yang untuk diambil dan permasalahan resultan mempunyai sampling. Moller (1982b, p.37): test kecakapan menyangkut alam semesta.

Penetapan validitas isi adalah characterising kecakapan bahasa dengan ketepatan cukup untuk memastikan contoh tugas tecakup dalam validitas ke perilaku kehidupan yang nyata dengan beberapa macam tugas tau metoda penilaian. Berbagai kesulitan membebaskan dicoba untuk membuat test relevan dalam kaitan dengan isi.

Prosedur merancang suatu test spesifikasi keterampilan boleh mendorong kearah variabilitas.

Suatu kebutuhan untuk meletakkan spesifikasi test untuk meyakinkan apa yang diuji. (Weir, 1983A) menyediakan informasi bermanfaat untuk menguji dan menerapkan spesifikasi dalam perwujudan test.

Pemeriksaan modul test bahasa oleh tenaga ahli tidak perlu untuk menjamin identifikasi proses benar-benar yang digunakan. Sebagai tambahan, akan bersifat berharga untuk mempekerjakan prosedur etnografi untuk menetapkan kebenaran materi.

Prosedur bermanfaat untuk populasi test introspek pada proses yang internal yang membawa penyelesaian materi test (lihat Aslanian), 1985; Cohen, 1985). Tenaga ahli akan berperan untuk penetapan petunjuk untuk melakukan prosedur metodologis jenis penyelidikan kebenaran test.

Test yang didasarkan pada keterampilan untuk menetapkan bahwa tingkat isi dibuat tegas/eksplisit. Porter (1983) menjeniskan validitas yang dilibatkan.

3.1.3 Face Validitas

Anastasi (1982, p.136):

Validitas menunjuk bukan apa yang benar-benar diukur, tetapi apa yang nampak diukur. Facen validitas menyinggung pada apakah test 'sah', siapa personil yang memutuskan penggunaannya.

Lado (1961), Davies (1965), E. Ingram (1977), Palmer (1981) dan Bachman dan Palmer (1981A) mempunyai nilai face validitas. Bachman dan Palmer (1981a, p.55) membantah sebagai berikut:

Karena tidak ada prosedur yang berlaku umum untuk menentukan ya atau tidaknya suatu test mempertunjukkan karakteristik, dan karena suatu basis yang bisa diterima untuk interpretative kesimpulan dari score test.

Jika suatu test tidak mempunyai face validitas, mungkin tidak bisa diterima oleh para siswa, atau para guru dan institusi. Jika para siswa tidak menerimanya sebagai sah, reaksi yang kurang baik mereka tidak melaksanakan dengan cara yang sungguh-sungguh mencerminkan kemampuan mereka. Anastasi (1982, p.136):

Jika isi test nampak tidak relevan, tidak sesuai, kekanak-kanakan atau pandir, hasil akan menjadi lemah, dengan mengabaikan validitas tidak cukup untuk suatu test secara objektif. Face validitas berfungsi secara efektif dalam situasi praktis.

Protes hukum empiris yang umum tentu saja diterapkan (Anastasi, 1982, p.136): 'Untuk memastikan validitas ditentukan...test validitas dalam format akhir dicek secara langsung.

Stevenson (1985b) menyatakan bahwa bentuk dan validitas isi harus ditingkatkan terhadap face validitas teknis.

3.1.4 Validitas *Washback*

Berbagai kesulitan menentukan apa yang perlu diukur lebih besar dengan apa telah baru-baru ini memasukakan 'adalah ITM-Morrow, 1986, atau lebih biasanya (Potter, 1983 dan Weir, 1983A) Washback test pada pengajaran dan pelajaran yang mendahului.

Para guru bahasa secara normal mencoba untuk memperlengkapi para siswa dengan keterampilan yang relevan untuk kebutuhan masa depan, dan tingkat test yang dirancang untuk mencerminkan hubungan antara test dan pengajaran.

Suatu pendekatan komunikatif pengajaran bahasa lebih mungkin diadopsi jika intruksinya sendiri komunikatif. Test bisa merupakan suatu instrumen yang

sangat kuat untuk mengakibatkan perubahan dalam kurikulum bahasa sebagai pengembangan terbaru dalam test bahasa. (Srilanka).

Ukuran untuk test komunikatif mungkin menjadi baik apabila mencukupi para siswa, para guru dan para pemakai hasil percobaan usaha sistematis untuk mengumpulkan data dapat dihitung pada validitas berdasar purbasangka validitas ia bermanfaat untuk menetapkan validitas terhadap ukuran-ukuran eksternal, sampai confirmatory suatu analisa statistik. Langkah pertama, dengan penekanannya pada bentuk, isi, wajah dan validitas washback, kemudian jenis test yang tersedia untuk eksternal prosedur dimaksudkan.

Karena bentuk, isi, wajah dan validitas washback adalah rumit. Ada suatu validitas yang terkait dengan ukuran dimana ukuran test tidak demikian rumit.

3.1.5 Validitas Terkait dengan Ukuran

Konsep kuantitatif, terkait dengan tingkat score test yang berhubungan dengan suatu ukuran pencapaian eksternal: Ingram (1977, p.18) 'validitas pragmatis'. Validitas terkait dengan ukuran membagi dalam dua jenis (Davies, 1977), validitas berbarengan dimana score test dihubungkan dengan ukuran capaian lain yang dibentuk pada waktu yang sama (Davies, 1983) dan validitas bersifat prediksi, dimana score test dihubungkan dengan beberapa ukuran capaian masa depan (Bachman dan Palmer, 1981A).

Pengesahan eksternal berdasarkan pada data selalu daripada 'armc spekulasi validitas isi'. Davies (1983, p.1) membantah bahwa pengesahan eksternal berdasarkan pada data selalu untuk lebih disukai: ukuran eksternal susah untuk ditemukan dan dilaksanakan dan mengukur suatu validitas test. Reliabilitas dan validitas internal sangatlah terikat. Anastasi : kebutuhan akan data eksternal dikumpulkan: 'Analisa internal test melalui korelasi item test, analisa materi test faktorial, dll. Tidak pernah cukup untuk pengesahan eksternal.

Konsep validitas terkait dengan ukuran menjadi empiricist-operationalist karena bahwa test dapat sah dengan cara tanpa perlu mengetahui apa yang test dapat sah dengan cara tanpa perlu mengetahui apa yang test sedang diukur, yaitu ketika ada bentuk validitas.

Morrow (1979, p.147)

Satuan asumsi tentang sifat alami bahasa dan pelajaran bahasa akan didorong ke arah test bahasa dengan sempurna dalam kaitan dengan asumsi ini tetapi nilai harus tak bias diacuhkan disangsikan jika asumsi yang basisi dasar diri mereka ditantang.

Jakobovits (1970, p.75): 'pertanyaan adalah untuk mengetahui suatu bahasa tidak baik dipahami dan, sebagai konsekuensi, kecakapan bahasa yang digunakan adalah tidak cukup sebab mereka mencoba untuk mengukurses suatu yang belum dirumuskan dengan baik'.

Test komunikatif menetapkan ukuran dibanding untuk menghubungkan itu. Hawkey (1982, p.153) meragukan test dalam suatu paradigma komunikatif: 'pada langkah pengembangan test komunikatif tersedia ukuran-ukuran untuk pengesahan berbarengan nampaknya kan lebih integrative/communicative dalam bentuk dan format dan untuk perbandingan langsung'.

Perbedaan boleh jadi dipaksa untuk menempatkan iman seseorang dalam suatu ukuran tetapi membangun yang dimasalahkan. Seseorang tidak bisa mengakui bahwa suatu test mempunyai validitas terkait dengan ukuran sebab berhubungan dengan test lain.

Test yang dipahami dalam paradigma komunikatif berbeda dengan hasil paradigma lebih awal. Penetapan ukuran-ukuran untuk mensahihkan test komunikatif bisa meragukan.

Penafsiran ukuran validitas terkait dianggap sebagai konsep berpotensi bermanfaat. Sebagai contoh, seseorang boleh jadi sangat hati-hati memproduksi hasil yang berlawanan dengan test yang mengukur ciri yang sama, terutama jika ditemukan untuk mempunyai bentuk validitas.

Validitas bersifat prediksi, mungkin saja bahwa dalam keadaan tertentu bersifat prediksi. Jika semua kekurangan adalah untuk memastikan ramalan tentang capaian masa depan atas dasar hasil percobaan, kekuatan ini memerlukan suatu test yang berbeda dengan menyediakan perbaikan. Jika ramalan dibuat atas dasar test akurat kemudian sifat alami test.

Validitas dan perkiraan keandalan berdasar pada correlation data harus diperlakukan dengan perhatian. Suatu korelasi tinggi boleh menandai adanya pengukuran dua atribut berbeda yang dihubungkan di antara populasi penempuh ujian. Pada sisi lain, korelasi rendah boleh menunjukkan dua atribut sungguh berbeda tertukar atau mencerminkan perbedaan kesalahan yang lebih tinggi dalam satu atau kedua test itu.

3.1.6 Bagaimana Seharusnya Suatu Test Dikenal?

Test GCSE dan test kecakapan bahasa, e.g., Universitas Cambridge (UCLES) Sertifikat kecakap dalam bahasa Inggris (CPE) dan Matrikulasi (Jt.1B) uji bahasa Inggris (luar negeri), bentuk operasional diatas tertarik akan kebenaran bersifat prediksi atau berbarengan, sedangkan, Davies (1982) telah menunjukkan berbagai hal ttg perhatian utama untuk EFL test tertutup paling distandardisasi. Menghubungkan hasil satu pengujian dengan yang lain atau terhadap beberapa ukuran masa depan dipandang sebagai suatu latihan tanpa makna ketika suatu baru pengujian baru telah dalam mempersiapkan tahun yang berikut dan hasil telah mengeluarkan untuk calon sekarang.

Test tertutup dihubungkan bahasa Inggris untuk tujuan bidang pendidikan, TEEP test (lihat catatan tambahan I), (ELTS) (lihat catatan tambahan V), rasa yang berkewajiban untuk melakukan diri prosedur pengesahan. Pengujian dilaksanakan tiap-tiap tahun untuk mempercayakan pada bentuk (tidak statistik), isi dan face validitas.

Dalam situasi dimana test adalah untuk suatu fungsi diagnostik derajat tinggi kejelasan didasarkan purbasangka langkah konstruksi test dirasakan untuk menjadi perlu. Tujuannya adalah untuk menyediakan statemen penuh arti pada suatu capaian yang akan berguna bagi mereka yang menyediakan pendukung mengenai perbaikan untuk dengan berbagai kesulitan.

Jika perhatian adalah untuk mengumpulkan informasi sesuai pada suatu capaian untuk kepentingan profil laporan bukannya untuk menetapkan suatu validitas test bersifat prediksi, tetapi lebih kewajiban untuk meningkatkan

content/construct validitas (itu) test dengan mengidentifikasi, konstruksi test, tugas komunikatif.

Usaha berdasar purbasangka untuk membuat test sama sah untuk menetapkan validitas test terhadap ukuran-ukuran eksternal. Jika langkah yang pertama dengan penekanannya pada validitas isi kemudian jenis test yang tersedia untuk prosedur pengesahan eksternal dimana test dimaksudkan.

Untuk menggambarkan kebangkitan baru minat berdasar purbasangka pengesahan test bisa jadi berguna. Mari kita mengasumsikan tugas adalah untuk membentuk suatu kecakapan test bahasa Inggris untuk Tujuan Akademis (Eap) yang menyediakan beberapa informasi diagnostik pada keterampilan studi yang terkait dengan bahasa.

Suatu test ttg materi bersifat ketatabahasaan bentuk terpisah untuk menghubungkan ukuran eksternal, sebagai contoh, test yang secara bersamaan mengaturukuran pada suatu tanggal, seperti test kenaikan kelas. Karena skornya rendah diadakan perbaikan. Score tunggal memerlukan informasi tentang gaya studi tertentu dimana seorang siswa mempunyai kesukaran mengoperasikan. Mereka mungkin secara lebih baik dilayani dengan test yang memperlihatkan bentuk, isi, dan face validitas. Orang tidak akan mampu mengalokasikan para siswa capaian dalam suatu discrete-point structuralist test yang kekurangan validitas ini.

Berdasarkan purbasangka pengesahan dari EAP test kecakapan dengan potensi diagnostik akan nampak untuk menuntut test terintegrasi macro-skills dibanding micro-elements. Jika tujuan adalah untuk menguji kemampuan yang komunikatif dari para siswa luar negeri, EAP adalah ragu-ragu apakah test kemampuan ilmu bahasa sendiri sebab bentuk test seperti itu perlu didasarkan pada tingkatan ilmu bahasa terpisah. Karena intisari komunikasi adalah suatu kemampuan untuk berkombinasi unsur-unsur ilmu bahasa terpisah dalam konteks tertentu,, kemampuan harus ditaksir dengan test keterampilan terintegrasi dibanding dengan test tingkatan ilmu bahasa terpisah.

Test EAP berdasar pada materi ilmu bahasa terpisah. EAP akan menjadi suatu sampling tugas yang komunikatif memerlukan para siswa di English-

medium. Kemampuan ilmu bahasa suatu test dipertimbangkan sah jika isinya didasarkan pada contoh khas unsur-unsur ilmu bahasa terpisah.

Menurut Canale dan Swain (1980, p.34) test komunikatif harus memahami tentang bahasa yang kedua dan bagaimana cara menggunakan (kemampuan) tetapi juga mempertunjukkan suatu situasi komunikatif penuh arti.

Penguji dipengaruhi apa yang Moller (1981b) uraikan yaitu paradigma sociolinguistik-communicative. Dengan singkat suatu test didalam paradigma komunikatif diharapkan untuk memperlihatkan corak yang berikut:

“akan ada suatu penekanan pada interaksi antar peserta, dan resultan intrasubjektivas akan menentukan bagaimana pertemuan ditingkatkan dan diakhiri”.

“format dan isi bahasa yang diproduksi adalah luas sehingga tak dapat diramalkan”.

“tujuan untuk memnuhi beberapa fungsi komunikatif”.

“mempekerjakan domain yang relevant dengan teks dan tugas asli. Kemampuan akan ditaksir dan mengembangkan konteks dan suatu profil capaian yang tersedia”.

“mengintegrasikan yang empat keterampilan membaca, mendengarkan, pidato dan menulis. Bahasa digunakan untuk ungkapan maksud fungsional; mempunyai arti penting”.

“menggunakan langsung metoda test, dengan mencerminkan proses realistik.”

“penilaian kemampuan produktif akan menjadi kualitatif bukannya kuantitatif, menyertakan penggunaan skala pengharkatan yang berkenaan dengan kategori capaian”.

Transaksi menguntungkan lebih harus ditunjukkan pada isi dan face validitas dibanding sebelumnya. Bagaimanapun, patokan yang bersifat elementer dalam pendekatan komunikatif ketest bahasa, akan bersifat bijaksana untuk sejumlah komponen ilmu bahasa utama.

Moller (1981b, p.44)

Pengujian komunikatif menguji aspek kecapakan tertentu. Tetapi adalah penting untuk menguji kecapakan bahasa tidak hanya untuk pengujian komunikatif. Capaian bahasa komunikatif dengan jrlas suatu dimensi kecapakan bahasa. Tetapi kemampuan bahasa adalah suatu dimensi kecapakan bahasa dan tidak bisa diabaikan. Juga harus diuji satu atau lebih orang banyak caya yang telah diteliti. Pengabaian dimensi adalah pengabaian re-awakening bahasa tradisional yang test komunikatif.

Revisi UCLES ELTS 1986-1989, IELTS (lihat catatan tambahan V), merencanakan test kosa kata dan tatabahasa dalam komponen yang umum. Dalam percobaan awal TEEP 1979-1982 adalah test pilihan item tatabahasa terpisah yang membuktikan indikator yang sah dan sempurna ttg kecapakan bahasa umum. Teep Riset menjelaskan bahwa komponen tatabahasa tidak menambahkan informasi kepada gambaran suatu kemampuan bahasa komunikatif. Karena pertimbangan serupa test kosa kata dan tatabahasa telah diberlakukan oleh IELTS.

Sejauh ini kita sudah berkonsentrasi pada cara test untuk meningkatkan validitas test dan melainkan fakta yang rumit kecuali jika test adalah tidak sah. Kebutuhan akan keandalan dalam rangka menjamin validitas test.

3.2 Konsep Reliabilitas

Suatu ukuran pokok dibanding dengan test bahasa manapun (Anastasi 1982; Guilford, 1965). Perhatian ada kemampuan tergantung pada hasil test dengan kata lain hasil diproduksi secara konsisten.

Tiga aspek reliabilitas pada umumnya diperhitungkan, pertama konsistensi membuat angka yang berbeda, sebagai contoh, ketika suatu test ttg ungkapan. Derajat tingkat inter-marker reliabilitas dibentuk dengan menghubungkan score yang diperoleh dengan lainnya. Konsistensi dari tiap inter-marker reliabilitas dibentuk dengan cara mereka untuk menghubungkan tanda pada kedua kesempatan (Anastasi, 1982).

Ukuran-ukuran penilaian perlu untuk dibentuk dan disetujui dan kemudian penanda perlu untuk dilatih; aplikasi ukuran-ukuran ini melalui prosedur standardisasi (I Murphy, 1979). Sepanjang tanda-tanda catatan ada perlu untuk suatu derajat tingkat cross-checking untuk memastikan bahwa standard disetujui.

Pertimbangan diperlukan untuk mencoba dan memastikan baha relevan sub-test internal konsisten. Semua materi dalam suatu sub-test ditentukan untuk mengukur atribut yang sama. Kuder-Richardson: untuk menaksir konsistensi internal disediakan manual statistik (lihat Anastasi, 1982, pp.114-6).

Aspek reliabilitas yang ketiga adalah bentuk reliabilitas paralel, kebutuhan dimana harus diingat ketika format alternatif masa depan suatu tet harus dipikirkan adalah sering sangat sukar untuk mencapai pertimbangan praktis dan teoritis. Untuk mecapai itu, dua versi alternatif suatu test perlu diproduksi. Reliabilitas versi berbanding lurus daiatur pada populasi test yang sama. Lebih sedikit reliabilitas dicek dengan metoda test-retest diamana tes yang sam diatur kembali pada populasi contoh yang sama.

Konsep reliabilitas penting ketika mempertimbangkan test bahasa dalam paradigma yang komunikatif (Poter, 1983). Davies (1965, p.14) menekankan: "Reliabilitas adalah penting bagi test; hanyalah untuk tertentu macam bahasa test tertentu yang mungkin sangat sukar untuk dicapai."

3.3 Validitas dan Reliabilitas Suatu Tegangan Tak Bisa Diacuhkan?

Pembatasan yang normal mempengaruhi pengembangan test terutama test prestasi dalam kelas. Perhatian pada umumnya berpusat pada pengesahan di langkah konstruksi test dan lebih sedikit luas dengan suatu posteriori pengesahan dilangkah capaian. Sumber daya untuk lakukan validitas bersifat prediaksi dan berbarengan, seperti yang diselenggarakan oleh Moller (1982b) dan oleh Institut Bahasa Universitas Edinburg, pada ELTS, tidaklah secara normal tersedia.

Perhatian dengan isi, bentuk dan valiitas test yang bersifat prediksi selalu diuji. Pengesahan membuktikan untuk menjadi usaha steril, kecuali jika kepedlian telah diambil alih reliabilitas test.

Masalahnya adalah bahwa seseorang dapat mempunyai reliabilitas test tanpa validitas test jika test dapat dipercaya. Kadang-kadang dikatakan menjadi suatu ketegangan reliability-validitas (Guilford, 1965 dan Davies, 1978). Ketegangan ini ada untuk mengorbankan suatu derajat tingkat reliabilitas peningkatan test yang mana suatu ukuran dapat dipercaya selain dari apa yang ingin diukur. Dua konsep dalam keadaan tertentu, satu sama lain eksklusif, tetapi jika suatu pilihan harus dibuat validitas adalah semakin penting. (Guilford, 1965, p.481).

Rea (1978) beryang argumentasi bahwa sebab test menilai bahasa komunikasi tidak bisa secara otomatis mengakui standard reliabilitas tinggi dengan cara yang sama discrete-item tidak diterima sebagai suatu pertimbangan untuk dilanjutkan pada ukuran yang sangat dapat dipercaya mempunyai; pertama mencoba untuk memperoleh ukuran yang dapat dipercaya ttg kemampuan komunikatif. Ini nampak lebih sedikit lebih masuk akal dan ekstrim dibanding yang diadopsi oleh Morrow (1979, p.151), membantah bahwa: Reliabilitas akan menjadi subordinat untuk menghadapi validitas. Obyektifitas palsu akan tidak lagi jadi suatu pertimbangan utama.

Sudut pandang Rea (1981a, p.x-xi), melaporkan bahwa pada April 1930 RELC seminar tentang Evaluasi dan Pengukuran Kemampuan bahasa dan Capaian, memutuskan hubungan adalah sangat dibutuhkan jika kita mengembangkan test prosedur yang secara sah mencerminkan pemahaman sifat alami kecakapan bahasa dan tujuan mengajar bahasa.

Rea menekankan bahwa: "suatu pendekatan pre-specific. Bahwa suatu tingkatan reliabilitas yang verifiable substansil harus pula dicapai, jika hasil percobaan adalah untuk mempunyai maksud." Moller mengadopsi suatu pendekatan serupa (1981a, p.67).

Bila dipahami bahwa suatu test sah harus dapat dipercaya dan akan nampak bahwa perilaku pribadi dan sangat kompleks seperti penggunaan dari bahasa ibu seseorang, validitas bisa diklaim untuk ukuran yang mungkin mempunyai lebih rendah bisa diterima tingkat reliabilitasnya.

Ia berargumentasi bahwa, walupun reliabilitas adalah sesuatu kita perlu selalu dicoba untuk mencapai test, mungkin tidak selalu menjadi pertimbangan yang utama dan menawarkan suatu posisi kompromi (p.67).

Dalam membentuk test yang berisi jenis tugas yang berbeda, sebagai contoh diperlukan untuk memperlihatkan suatu derajat reliabilitas tinggi. Terutama sekali test ttg penggunaan komunikatif, memperlihatkan reliabilitas lebih rendah tanpa dengan kurang baik mempengaruhi keseluruhan validitas.

Hawkey (1982,p.149):

Reliabilitas suatu test tidak bisa diabaikan tanpa suatu efek berbahaya pada validitas instrumen. Tetapi ada kemungkinan bahwa, jika bentuk validitas ttg test komunikatif untuk dipastikan, reliabilitas pertanyaan akan harus diterima sebagai subordinat.

Validitas adalah penting sebab dihubungkan dengan cara yang ditempuh oleh capaian test. Houston (1983) menguraikan perbedaan antara norma dan metoda criterion referenced yang melukiskan tingkatan dan mendiskusikan sebagai dari berbagai kesulitan ukuran-ukuran capaian sesuai metode yang belakangan dipilih.

Popham (1978, p.2)

Suatu test criterion-referenced dirancang untuk menghasilkan suatu uraian jelas bersih dari apa yang dicapai. Dibanding menginterpretasikan test penempuh ujian dalam hubungan pada capaian test tradisional, test criterion-referenced baik menghasilkan suatu gambaran yang lebih baik.

Davies (1978, p. 158) yang dibuat koneksi dengan pengujian bahasa dan menyatakan reservasi tertentu tentang test criterion-referenced.

Ada berbagai kesulitan dalam menggunakan test criterion-referenced untuk bahasa: tidak ada inventori poin-poin atau materi terbatas; ada sasaran perilaku; ada variabel ukuran-ukuran sukses eksternal, kelancaran, kejelasan, dll; tidak ada cara yang jelas nyata menetapkan pengetahuan cukup, tentang suatu bahasa.

Dengan begitu sebagian dari berbagai kesulitan menunjuk kemudiannya oleh Houston (1983) ditaruh adalah suatu bahasa yang menguji konteks, [yang] dengan jelas, criterion-referenced tingkatan capaian adalah mungkin hanya untuk tingkat [bahwa/yang] test mempunyai suatu derajat tinggi kebenaran isi.

3.4 Uji Efisiensi

Suatu test dapat dipercaya dan sah jika membuktikan kepraktisan: melibatkan pertanyaan ekonomi, administrasi, membuat angka, dan penafsiran hasil.

Lamanya percobaan untuk mempengaruhi suksesnya cara-cara lain, sebagai contoh., suatu kelelahan mempengaruhi calon, faktor administratif seperti yang mengawas ujian dan ketersediaan tempat dimana untuk duduk.

Untuk menyediakan profil skor baku, masing-masing bagian dari profil yang harus dijangkau suatu derajat tingkat keandalan bisa diterima. Untuk mencapai keandalan memuaskan, test komunikatif mungkin mempunyai angka. Berbagai kesulitan untuk menjamin bahwa test berisi suatu contoh yang mewakili tugas. Untuk tingkatan kebenaran dengan menyediakan makanan kebutuhan spesifik dan bagan, lebih banyak test diperlukan biaya-biaya meningkat dibandingkan dengan mereka yang test [yang] umum tunggal tersedia untuk populasi besar.

Efisiensi kelangsungan hidup keuangan, boleh membuktikan dalam pengembangan ttg test komunikatif. Test jenis ini adalah sulit dan waktu yang dikonsumsi untuk membentuk lebih lama, sumber daya untuk mengurus menuntut standardisasi dan pelatihan pemeriksa saksama menjadi lebih mahal. Ongkos penggunaan test komunikatif dalam pengujian besar-besaran operasi sungguh membatasi penggunaan mereka.

Suatu kebutuhan sangat mendesak untuk mencoba dan mengembangkan format test dan ukuran-ukuran evaluasi yang menyediakan keseluruhan antara keadaan, kebenaran dan efisiensi dalam penilaian ttg keterampilan komunikatif.

Dalam bab tiga kita kembali ke perhatian lebih praktis dan langkah-langkah dalam pengembangan suatu test untuk memberi suatu gagasan proses secara normal diikuti disain dan implementasi suatu bahasa.