

BAB II

PENDEKATAN TES BAHASA

1.1 Pendahuluan

Dalam suatu tes diperlukan sebuah pendekatan. Pentingnya sebuah alternatif pendekatan tes bahasa berkaitan dengan ukuran-ukuran validitas, reliabilitas dan efisiensi.

Validitas dikaitkan dengan apakah tes mengukur apa yang hendak diukur. Reliabilitas berkaitan dengan tingkatan yang mana pada hasil tes. Efisiensi berkaitan dengan berbagai hal pelaksanaan dan desain tes dan pengadministrasiannya. Rangkaian ini harus cukup untuk ditinjau ulang dalam tes bahasa.

Daves (1978) berargumentasi bahwa pada pertengahan 1970-an, pendekatan terhadap test nampak lepas dari rangkaian 'terpisah', seperti *cloze test* dan lainnya. Ia berpendapat bahwa dalam pelaksanaan tes, ada suatu permasalahan antara analitis yang disampaikan dan keintegralan pada lainnya. Selain itu, pertimbangan pandangan tes bahasa yang memuaskan dan macam tes bahasa yang paling bermanfaat adalah suatu kombinasi dua pandangan (analitis dan integral). Tetapi, mungkin ada tes biasa secara keseluruhan integratif dan analitis.

Kedua hal ini (analisis dan pengintegrasian) memiliki kemiripan dan hubungan erat terhadap konsep validitas dan reliabilitas. Validitas tes meningkat dengan menambah jumlah materi yang terpisah dalam suatu test: semakin sedikit semakin kecil, semakin banyak semakin tinggi tingkat reliabilitasnya. Validitas sebuah tes akan meningkat dengan membuat tes yang lebih nyata dalam kehidupan dan penggunaan bahasa.

Pada sisi lain, Oller (1979) merasakan bahwa tes perlu memusatkan pada akhir rangkaian integratif. Ia membuat suatu kritikan kuat terhadap Spolsky (1976) yang telah diuraikan pada '*Era psychometric-structuralis*', atau yang disebut 'titik terpisah' tentang tes, ke apa yang ia memasukkan dalam '*Era psycholinguistic-sociolinguistik*': tentang integratif.

Uraian tentang pendekatan sebuah tes bahasa, siswa diperlukan sebagai objek dalam praktek, kebanyakan test berisi unsur-unsur yang terpisah dan integratif, baik format test atau prosedur penilaian.

1.2 Era Psychometric-Structuralis

Keuntungan yang jelas dari test ‘terpisah’ pon-poin ilmu bahasa adalah bahwa mereka menghasilkan data yang mudah dapat dihitung, seperti halnya membiarkan suatu pemenuhan materi yang luas. Tes ‘terpisah’ yang memusatkan materi ilmu bahasa adalah efisien dan mempunyai keandalan yang umum dengan test yang dicapai, tetapi kedua-duanya pendekatan dan format yang digunakan memiliki kelemahan untuk mengukur.

Analisis poin terpisah memerlukan pemecahan unsur-unsur bahasa dan mencoba untuk mengerjakan atau menguji secara terpisah dengan sedikit atau tidak ada perhatian terhadap cara unsur-unsur itu saling berhubungan merupakan suatu konteks komunikasi yang lebih besar. Apa yang membuat tidak efektif sebagai basis untuk mengajar atau menguji bahasa adalah bahwa kekayaan bahasa yang rumit hilang ketika unsur-unsurnya dipisahkan. Fakta adalah bahwa sistem manapun jika bagian yang saling berhubungan menghasilkan kekayaan dan kualitas jika tidak ada bagian yang terpisah.

Pandangan Oller ini merupakan landasan yang aman dan kebanyakan orang mungkin akan setuju bahwa menguji kemampuan calon ilmu bahasa adalah suatu yang diperlukan, tetapi tidak cukup, komponen test. Dalam konteks lain, pemilihan seorang tester diperlukan untuk mempertunjukkan bahwa mereka dapat melaksanakan tugas. Otoritas tidak tergantung semata-mata pada cara menullis dan catatan. Dengan cara yang sama, mereka yang harus membuat penilaian tentang suatu potongan bahasa yang mereka buat dari potongan secara keseluruhan.

Chaplen (1970) mengkritik keterampilan terisolasi yang menguji dari segi pandangan ini, bahwa: mau tidak mau penilaian itu merupakan keterampilan komponen yang paling umum terisolasi dengan menyediakan satu demi satau atau dalam bentuk kumpulan. Gestalt memiliki pandangan yang sama dengan Saviagon

(1972) yaitu menemukan kemampuan yang bersifat tatabahasa itu tidaklah tentang keterampilan komunikatif.

Kelly (1978) yang berargumentasi bahwa jika tujuan dari linguistik diterapkan dilihat ketika analisis yang maksud diterapkan, sebagai contoh, pengenalan maksud context-specific dari suatu ucapan sebaliknya dari yang system-giving maksud, kemudian alih bahasa yang diterapkan harus lebih menarik untuk pengukuran dan pengembangan kemampuan untuk menetapkan capaian komunikatif, produksi dan pengertian tentang ceramah terpadu, dibanding dalam kemampuan ilmu bahasa. Spolsky's (1968) menitikberatkan pada awal untuk mencoba menetapkan pengetahuan bahasa seseorang dalam kaitan dengan suatu penguasaan persentase tatabahasa dan kosa kata, kita akan lebih baik dihadapkan pada test untuk melaksanakan suatu sosiolinguistik.

Rea (1978, p.51) berpandangan serupa “Walapun kita akan setuju bahasa itu adalah suatu perilaku kompleks dan bahwa kita biasanya akan menerima suatu definisi keseluruhan kecakapan bahasa sebagai kemampuan untuk bahasa alami, atau membiarkan orang lain memaksakan kita ukuran untuk menilai bahasa sebagai suatu yang abstrak dari materi terpisah, untuk digerakan hanya pada suatu cara mekanistik. Test seperti itu menghasilkan jenis tiruan, tidak relevan dan steril yang materi tidak punya hubungan terhadap penggunaan bahasa dalam situasi hidup riil”.

Morrow (1979) menjelaskan bahwa jika kita menilai kecakapan, yaitu sukses potensial dalam penggunaan bahasa dalam beberapa pengertian umum, itu akan lebih berharga menguji suatu pengetahuan dan suatu kemampuan menerapkan aturan dan proses dengan mana unsur-unsur yang terpisah ini disatukan kedalam suatu tanpa batas jumlah kalimat bersifat tatabahasa dan kemudian memilih sebagai hal konteks tertentu, bukannya hanya menguji pengetahuan unsur-unsur.

Morrow (1979) membantah bahwa, “pengetahuan unsur-unsur suatu bahasa sesungguhnya sama sekali tidak berarti kecuali jika pemakai bisa mengkombinasi dengan sesuai dan menemukan situasi bahasa di mana ia mengharapkan menggunakan bahasa itu”.

1.3 Era Psycholinguistic-Sociolinguistic

Karena jawaban atas suatu test ‘terpisah’ adalah indikator kecakapan bahasa tidak cukup, menguji keseluruhan pada test global pada 1970-an, dalam Spolsky (1976) era psycholinguistic-sociolinguistik, pendekatan ke arah pengukuran.

Oller (1979) mengklaim bahwa test integratif global seperti cloze dan dikte melampaui pengukuran bagian dari terbatas kemampuan bahasa terbatas yang dicapai oleh test terpisah dengan penyimpangan mereka kearah test keterampilan yang diterima; bahwa, test seperti itu bisa mengukur kemampuan mengintegrasikan keterampilan bahasa berlainan dalam cara-cara yang mana lebih lekat mendekati proses penggunaan bahasa yang nyata.

Konsep dari suatu test integratif telah dilahirkan untuk membandingkan definisi suatu test terpisah. Jika materi terpisah menjarangkan; memisahkan keterampilan bahasa, test integratif menaruh kembali bersama-sama. Sedangkan materi terpisah mencoba untuk menguji pengetahuan bahasa lebih sedikit pada waktu yang sama, tes integratif mencoba untuk menilai suatu kapasitas pelajar untuk menggunakan lebih banyak pada waktu yang sama, dan mungkin beberapa komponen sistem yang bersifat tatabahasa, dan barangkali lebih dari satu aspek atau keterampilan yang secara kebiasaan mengarang.

Dari perspektif psycholinguistik, bahasa dilihat menjadi seperti lebih sedikit memiliki suatu struktur taxonomi yang dirumuskan dengan baik dan lebih banyak suatu sistem fungsional yang kreatif. Sehingga menimbulkan suatu pemborosan, sedemikian sukar untuk menunjukkan bahwa semua unit ilmu bahasa tunggal adalah sangat dibutuhkan untuk komunikasi. Kontribusi yang sociolinguistic berpusat pada konsep tentang kemampuan komunikatif yang mengahdirkan suatu perluasan dugaan. Chomsky berpendapat tidak hanya pengetahuan aturan untuk membentuk kalimat yang bersifat tatabahasa tetapi juga aturan untuk menggunakan kalimat itu yang sewajarnya dengan konteks berbeda. Dengan begitu perspektif sociolinguistik dan psykolinguistik sudah memperbesar

basis yang mendasari kebenaran suatu test. Ukuran-ukuran baru sudah diperkenalkan yang tidak bisa diukur dengan metode standar sasaran.

Oller menganggap ilmu bahasa memerlukan capaian dibawah batasan hidup riil, sebagai contoh, waktu merupakan suatu pemandu keserasian dan potensi komunikasi, sekalipun mereka menguji komunikasi.

Alderson (1978) berpendapat bahwa mengangkat serius pertanyaan tentang kebenaran ukuran yang integratif ini sebagai alat test. Ia mempertunjukkan bahwa tidak ada cloze test dan bahkan dalam menggunakan jalan lintas yang sama, hasil Cloze Producer diuji secara lebih detail dalam bagian 4.1.3 di mana pertimbangan diberikan kepada konstruksi dan keuntungan yang potensiak dan kerugian-kerugian dari metode test ini.

Penyebab perhatian utama adalah asumsi yang dibuat oleh Oller (1976, 1979, 1980) bahwa kecakapan bahasa umum (Glp), tatabahasa test integratifnya diarahkan pada, adalah faktor utama tunggal yang mendasari semua keterampilan bahasa. Konsepnya keseluruhan kecakapan telah tak bisa diacuk\hkan dalam suatu hipotesis dari suatu dasar kemampuan.

Pandangan Oller tentang konsep tatabahasa yang internalized adalah sesuatu yang menggiurkan untuk epentingan mereka mengambil keputusan administratif, Davies (1981) poin-poin itu bertentangan dengan bukti substansiil yang berpihak pada setidaknya dua kemampuan, yakni resepsi dan produksi (lihat Vollmer, 1981). Perbedaan antara pengetahuan bagaimana cara meneliti masukan dan mengetahui bagaimana cara membangun keluaran akan nampak lebih memberatkan dibanding kedua proses. Pengalaman yang bersifat pendidikan akan juga menyatakan bahwa tugas pencapaian perorangan yang berbeda mengakibatkan berbagai kecakapan berbeda.

Davies (1981) menekankan bahwa test integratif mengahdirkan total kecakapan bahasalebih baik daripada kombinasi atau test tunggal, suatu argumentasi tidak dengan sendirinya berpihak pada hipotesis kesatuan kemampuan. Cloze dan dikte menjadi sangat integratif karena mereka mengisi paling banyak atau semua kemampuan bahsa. Korelasi tinggi antara cloze dan ukuran lain hanya mencerminkan bahwa mereka sedang mengukur. Perbedaan

keterampilan adalah sangat dihubungkan antar individu, ini tidak berarti bahwa tidak akan ada pencapaian individu dalam berbagai perbedaan keterampilan.

Kelompok penempuh test mungkin punya scor pada test dalam pemahaman yang kedua-duanya pada rangking yang sama, tetapi karena co relatoinal ukurannya sedikit atau tak berguna, score kelompok yang mungkin dipusatkan pada makna sangat berbeda pada keduanya, menandakan tingkat perbedaan pencapaian keseluruhan. Dengan kata lian, co relational data tidak menyediakan bukti tentang standar.

Keterangan empiris yang telah disusun perbihak pada hopotesisi kesatuan kemampuan terbuka terhadap beberapa keraguan adalah suatu kuti yang berkembang untuk menyokong hipotesis keadaan.

Komponen analisa utama adalah sering digunakan untuk memperkuat hipotesis kesatuan kemampuan, tetapi metoda ini sangat utama dirancang untuk menyederhannakan data. Dan akan diharapkan untuk menghasilkan satu faktor dari suatu test bahasa. Pada umumnya kecakapan bahasa tidak perlu menjelaskan semua perbedaan dlam hasil, dan persen dari perbedaan belajar. Oleh karena keberadaan faktor selain dari komponen yang prinsip, yang menjelaskan proporsi perbedaan yang layak adalah sering mungkin dengan pengejaran analisis faktor lebih lanjut. Sebagai contoh Varimax berpendapat tentang faktor struktur, untu memperoleh sejumlah fator mandiri hasil masing-masing merupakan suatu kontribusi cukup besar terhadap total perbedaan.

Ada juga bukti literatur bahwa format suatu tugas dapat terlalu mempengaruhi pencapaian ((lihat Boniakowska, 1986; Murphy, 1987, 1980: dan Bendungan , 1983A). bentuk tesebut diperlukan berbagai format test untuk menaksir masing-masing keseluruhan ukuran.

Meskipun test Oller telah memperlihatkan pada aspek yang berbeda dalam sistem bahasa, mereka tidak dapat secara langsung. Walaupun test tersebut mengintegrasikan keterampilan bahasa berlainan dalam cara-cara yang lebih lekat mendekati bahasa nyata yang digunakan, orang akan membantah bahwa klaim mereka terhadap kebenaran komunikatif, seperti hanya menirukan relevan dengan komunikasi nyata.

Advokasi tentang test bahasa komunikatif akan membantah bahwa pandangan Oller tidak cukup terhadap pentingnya pengolahan ceramah yang produktif, sebagai akibat dari penggunaan bahasa yang nyata dalam suatu konteks sosial dengan semua batasan capaian, sebagai contoh, dalam ceramah interaction-based, hasil perilaku dan kemungkinan tindakan. Kedua-duanya sudah menekankan bahwa walaupun ukuran kemampuan bahasa tidak langsung mengakui standar keandalan yang sangat tinggi dan kebenaran berbarengan terbentuk oleh teknik klaim mereka kepada lain jenis kebenaran.

Morrow (1979) memberikan fakta bahwa cloze maupun dikte menawarkan kesempatan untuk memproduksi secara spontan oleh calon dan norma-norma bahasa yang diikuti adalah dari pemeriksa, bukan siswa itu sendiri. Test prosedur tidak menawarkan kemungkinan untuk produksi tertulis atau lisan karena biasanya diadakan untuk yang sangat penting, makna beberapa menaksirkan untuk situasi komunikatif. Walaupun ukuran integratif nampak menghubungkan ukuran yang serupa dari kecakapan bahasa umum. Ada keterangan empiris pada cloze yang dihubungkan dengan test produksi tertulis. Test yang terkait adalah dapat dipercaya, dan menyatakan bahwa kemungkinan area kecakapan ini tidak bisa cukup diramalkan oleh suatu test keseluruhan kecakapan.

Morrow juga mengklaim kedua-duanya cloze dan dikte pada dasarnya adalah test dasar kemampuan bukannya pencapaian nyata. Dengan kata lain, mereka tergantung pada dasarnya pengetahuan sistem bahasa bukannya kemampuan untuk mengoperasikan sistem. B.J Carroll (1980b, p.9) menyimpulkan bahwa cloze test masih sangat utama pemakaian mendasar. Tugas tidak menghadirkan komunikasi interaktif asli dan oleh karena itu, komunikatif sehari-hari dapat mengatasi tasks.

Sekalipun diputuskan test tidak langsung seperti cloze yaitu jenis tugas pertunjukan yang diperoleh, jenis aktivitas membaca. Konsekwensi yang bersifat pendidikan mencakup test jenis ini adalah mengukur secara rinci untuk menangani tugas penilaian tidak langsung dalam bentuk pilihan yang lebih realistis.

Kelly (1978, p.24) berpendapat bahwa diberi tugas dengan pelatihan tertentu.

Siswa menganalisa jawaban suatu test tidak langsung tidak akan memberika informasi yang relevan menyangkut pertimbangan berbagai kesulitan siswa dalam tugas yang asli, test yang tidak langsung adalah suatu ukuran yang dapt dipercaya dan sah. Dengan test tidak langsung dapat memberikan bukti untuk tingkat pencapaian, tetapi tidak bisa mendiagnose area kesukaran yang spesifik dalam hubungan dengan tugas yang asli itu.

Test integratif seperti cloze hanya menceritakan sekitar kemampuan ilmu bahas. Mereka tidak menceritakan secara langsung sekitar kemampuan pencapaian siswa, dan nilai utama dalam format tidak dimodifikasi dalam bentuk kemampuan untuk mengukur bukannya hubungan pencapaian ukuran-ukuran eksternal. Mereka barangkali hanya penggunaan minat yang terbatas dengan apa yang dapat inividu atau tidak bisa dilakukan dalam kaitan dengan berbagai tugas bahasa yang mungkin ia dalam situasi hidup riil.

Defisiensi dalam jenis informasi pendekatan test terpisah era psychometric-structuralist dan pendekatan integratif era psycholinguistik-sosiolinguistic dapat menyediakan kesempurnaan suatu kebutuhan untuk menyellidiki paradigma yang komunikatif untuk melihat apakah pendekatan ini lebih memuaskan.

1.4 Paradigma Komunikatif

1.4.1 Istilah

Istilah dalam beberapa literatur pada pendekatan komunikatif terhadap bahasa acu pencapaian, sebagai contoh, Buku Carroll (1980b) berjudul Menguji Pencapaian Komunikatif. Menyatakan test pencapaian adalah pencapaian perorangan pada diri seseorang terhadap situasi, sekitar kemampuan untuk menangani situasi lain, kemampuan seperti halnya pencapaian adalah melibatkan dengan tepat kapasitas dalam pengertian (Widdowson, 1983). Bachman: istilah kemampuan bahasakomunikatif meliputi pengetahuan kedua-duanya, kemampuan

bahasa dan kemampuan untuk menerapkan kemampuan dalam bahasa. Widdowson: kemampuan dan definisi kecakapan bahasa memuaskan.

Pada hakekatnya, test pencapaian adalah satu perilaku dalam pengaturan tunggal dengan tidak ada niat menyamaratakan di luar ketentuan-test bahasa komunikatif harus dilakukannya sendiri kapasitas (Widdowson, 1983) atau kemampuan bahasa komunikatif (Bachman, 1990). Seluruh tindakan menyamaratakan di luar pengaturan yang benar-benar menguji kemampuan untuk menggunakan pengetahuan dan/atau bahasa tentangnya. Dan sebaliknya kemampuan itu untuk dilihat bagaimana kemampuan (mengetahui tentang penggunaan suatu bahasa) boleh jadi dievaluasi kecuali melalui perwujudannya dalam pencapaian. Pencapaian hanya dapat secara langsung diamati dan karenanya dievaluasi.

Dalam menguji kemampuan berbahasa komunikatif kita mengevaluasi contoh pencapaian, dalam konteks yang spesifik, diciptakan di bawah batasan test tertentu. Untuk apa mereka dapat menceritakan suatu kemampuan bahasa atau kapasitas calon komunikatif. Skehan (1988) berpendapat bahwa test tidak boleh replicate persis kondisi-kondisi suatu tugas secara spesifik dalam situasi target sampai taraf kondisi-kondisi tertentu tentang pencapaian nyata.

Apa yang kita butuhkan adalah suatu teori untuk memandu dan meramalkan bagaimana dasar kemampuan komunikatif dinyatakan sebagai pencapaian nyata: bagaimana situasi dihubungkan dengan satu sama lain, bagaimana kemampuan dapat ditaksirkan dengan contoh pencapaian tes nyata: komponen/kemampuan wewenang komunikatif apa yang benar-benar dipunyai; dan bagaimana hubung satu sama lain... karena teori itu pasti tidak ada, pengujian harus melaksanakan tugas mereka seperti teori yang tersedia.

1.4.2 Dasar Teoretis

Kevaliditasan tes secara komunikatif adalah suatu fungsi derajat tingkat pemahaman komunikasi dan kemampuan komunikatif pada pihak perancang test. Adalah naif untuk berasumsi bahwa seseorang dapat mengembangkan tes yang

sah tentang kemampuan bahasa komunikatif tanpa memandang merancang sesuatu sambil mencoba untuk mengukur, argumentasi yang berkenaan dengan status bahasa uraian tidak disediakan.

Persetujuan pada komponen apa yang harus tercakup disuatu model tentang kemampuan bahasa komunikatif tidak sama sekali dengan suara bulat (lihat Courhene, 1985, p.49). tentu saja relatif kecil diketahui paradigma komunikatif yang lebih luas jika dibandingkan dengan kemampuan ilmu bahasa yang didalam dirinya dan cukup dikembangkan teori tentang waktu penggunaan bahasa komunikatif belum tersedia. Ini berarti bukan harus menantikan penyelesaian. Teori sebelumnya dapat dikembangkan untuk menguji prosedur. Melainkan kita harus menyelidiki secara sistematis sebagian dari hipotesis yang tersedia tentang penggunaan bahasa dan mencoba mengoperasionalkannya untuk menguji tujuan. Dengan cara ini membangun dan proses ttg linguistik terapan mungkin diuji dengan pengalaman dan status mereka.

Canale dan Swain (1980) menyajikan klarifikasi istilah yang diperlukan untuk membentuk suatu kemampuan penggunaan bahasa secara komunikatif. Penulis berpendapat bahwa kemampuan komunikatif meliputi kemampuan bersifat tatabahasa (pengetahuan ketentuan-ketentuan tatabahasa), kemampuan sociolinguistic (pengetahuan ketentuan-ketentuan penggunaan dan peraturan tentang ceramah) dan kemampuan strategis (pengetahuan ttg strategi komunikasi tidak lisan dan lisan). Model itu dibaharui oleh Canale (1983), yang mengusulkan suatu forum-dimensional model yang berisikan ilmu bahasa, sociolinguistik, discoursal dan kemampuan strategis; pembedaan tambahan dibuat antara sociolinguistik (sociocultural) kemampuan (aturan) dan discoursal kemampuan (kohesi dan lekat).

Kemampuan bahasa terdiri atas keterampilan dan pengetahuan yang spesifik yang diperlukan untuk menetapkan maksud ucapan, untuk memanfaatkan bahasa yang sesuai terhadap konteks dan untuk mengoperasionalkan melalui bahasa di luar tingkatan kalimat. Kemampuan strategis terdiri dari keterampilan dan pengetahuan yang semakin umum melibatkan, menaksir, merencanakan dan melaksanakan tindakan komunikatif secara efisien. Skehan (1988) menyatakan

bahwa komponen strategis mencakup ketika komunikasi memerlukan penciptan mendadak sebab kemampuan lain adalah dalam beberapa cara tidak cukup. Bagian akhir dari model Bachman yaitu faktor keterampilan dan metoda untuk menangani operasi bahasa yang nyata dalam situasi riil dan demikian menempatkan kemampuan dalam kerangka capaian yang lebih luas.

Model seperti ini memberikan kerangka yang bermanfaat untuk perancangan test bahasa, tetapi harus ditekankan bahwa mereka masih diri mereka (lihat Brindley, 1986; Swan, 1985). Keberadaan komponen bahkan waktu kesatuan terpisah belum terbentuk. Skehan (1988) menunjuk bahwa hubungan antara berbagai kemampuan tidaklah seluruhnya jelas, maupun terintegrasi kedalam keseluruhan kemampuan komunikatif. Maupun apakah dijelaskan bagaimana kemampuan kominakif ini diterjemahkan kedalam pencapaian komunikatif. Candlin (1986) juga menguraikan sebagian dari permasalahan untuk dihadapkan kemampuan komunikatif dan berargumentasi bahwa solusi mereka tergantung pertama pada uraian kita ttg rancangan.

Sampai saat ini sejumlah riset telah dilaksanakan terbatas pada penyelidikan, pengukuran kemampuan bahasa dan pfaktor metoda tetapi sangat kecil telah dilaksanakan pada pengukuran strategi komunikasi yang spesipik atau hubungannya terhadap kemampuan lain. Ini dengan sendirinya mungkin merupakan indikasi yang tidak bisa dipisahkan dari berbagai kesulitan dalam area ini. Ada suatu kebutuhan mendesak untuk riset sistematis untuk menerangi semua isu yang belum terpecahkan ini.

Untuk membantu menjelaskan apa yang dimaksud dengan test komunikatif kita meminta pertolongan pretheoretical data yang tersedia dari literatur yang berkenaan dengan konsep tentang kemampuan komunikatif. Karena dua model dimensional Hymes tentang kemampuan komunikatif, berisikan unsur ilmu bahasa dan sociolinguistic, model yang berikut telah mempertimbangkan suatu yang dimensi sociolinguistic untuk mengenali pentingnya konteks terhadap penggunaan bahasa yang sesuai dan interaksi yang dinamis yang terjadi antara konteks dan ceramah.

Hymes (1972) berpendapat bahwa kemampuan komunikatif yang telah mencakup kemampuan untuk menggunakan bahasa, seperti halnya mempunyai pengetahuan yang mendasari pencapaian nyata. Morrow (1972): suatu perbedaan perlu dibuat antara kemampuan komunikatif dan pencapaian komunikatif, corak pembeda yang belakangan menjadi fakta bahwa pencapaian adalah perwujudan. Canale dan Swain's (1980): tiga kemampuan dan interaksi mereka.

Morrow (1979): pengertian dan produksi ucapan yang nyata (dibawah jenis/batasan psikologis adalah unik terhadap pencapaian).

Morrow (1979), Canale dan Swain (1980) berargumentasi bahwa bahasa komunikatif terkait dengan apa yang pelajar memahami tentang format bahasa dan sekitar bagaimana cara menggunakannya sewajarnya dalam konteks yang berguna (kemampuan), harus pula berhadapan dengan tingkat untuk pelajar yang mana yang benar-benar mampu mempertunjukkan pengetahuan dalam suatu situasi komunikatif yang penuh arti (pencapaian), yaitu, apa yang dapat ia lakukan dengan bahasa, atau seperti pendapat Era (1978, p.4): 'kemampuan berkomunikasi dengan kesenangan mempengaruhi penentuan sociolinguistic'.

Kapasitas atau kemampuan (lihat Widdowson, 1983; Bachman, 1990) untuk menggunakan bahasa secara komunikatif melibatkan kemampuan kedua-duanya kemampuan mendemonstrasikan dan menggunakan kemampuan bahasa. Tugas pencapaian dihadapkan dengan tes komunikatif harus terwakili; contoh jenis tugas yang mungkin dihadapkan situasi kehidupan nyata mereka sendiri dan sesuai dengan bahasa normal yang digunakan dimana situasi pengintegrasian ttg keterampilan komunikatif diperlukan dengan waktu yang sedikit untuk, atau memonitor masukan bahasa dan keluarannya. Ukuran-ukuran tugas penilaian pencapaian ini perlu dihubungkan lekat dengan komunikasi gagasan yang efektif dalam konteks.

Perspektif dikonsistenkan dengan tester bahasa yang biasanya mendukung suatu model dasar kemampuan bahasa komunikatif dimana penekanan pergeseran dari ilmu bahasa kepada dimensi yang komunikatif. Penekanan adalah tidak lagi pada ketelitian ilmu bahasa, tetapi pada kemampuan untuk memfungsikan secara efektif melalui bahasa dalam konteks situasi tertentu.

Cooper (1968) memandang kerangka tes ada itu, sebab mereka berkonsentrasi pada kemampuan ilmu bahasa benar-benar menggunakan bahasa untuk membaca, tulis, berbicara atau mendengarkan dalam cara-cara dan konteks sesuai dengan hidup riil.

B.J. Carroll (1989b, p.1): kebutuhan utama pelajar kebanyakan bukanlah suatu pengetahuan analitis atau teoritis target bahasa, tetapi kemampuan memahami dan dipahami yang bahasa dalam konteks dan batasan menggunakan bahasa dan keadaan tertentu.

Pendapatnya adalah bahwa: “penguasaan ukuran bahasa yang terakhir kemudian efektivitas pelajar dalam komunikasi untuk mengatur temuan dirinya sendiri”.

Era (1985) berargumentasi bahwa semua test dapat dilihat ketika test pencapaian merupakan macam-macam derajat tingkatan komunikatif atau tidak komunikatif, dasar penggunaan. Lebih lanjut membedakan antara materi sebagai meaning-dependent, dan menguraikan bagaimana yang terdahulu dapat dibagi lagi menurut keterlibatan suatu konteks untuk menentukan tanggapan atau bukan. Untuk mengambil bagian dalam suatu peristiwa komunikatif adalah dengan menghasilkan dan memahami ceramah dalam konteks situasi dan di bawah kondisi-kondisi pencapaian yang diperoleh. Tujuan test kecakapan untuk menilai ya tau tidaknya calon mampu mengambil bagian peristiwa komunikasi dari komunikasi yang diterapkan.

Pernyataan mencerminkan suatu penekanan bahasa yang diajarkan, test telah ditempatkan pada penggunaan dan perhatian yang telah ditunjukkan dengan fungsi komunikatif bukannya dengan pola pemakaian bahasa formal (lihat Campbell dan Wales, 1970; Hymes, 1972 dan Widdowson, 1987 dan 1983). Uraian teoritis adalah penting untuk menggambarkan parameter yang lebar dalam test bahasa komunikatif tetapi kebutuhan praktisi lebih terukur untuk memastikan derajat tingkat sifat, suka berbicara, suatu test atau untuk membuat test merekasama sekomunikatif mungkin dalam batasan yang diperoleh. Seperti apa suatu test kelihatan komunikatif? Bagaimana cara membedakan dengan test lain? Pertanyaan itu yang harus kita fahami.

1.4.3 Corak Pembeda tentang Menguji Bahasa Komunikatif

Hanya sedikit test bahasa yang menggunakan teori penggunaan bahasa. Maka keterampilan dan kondisi-kondisi pencapaian adalah penting untuk menilai kemampuan bahasa komunikatif. Test pembentukan harus lekat untuk mengidentifikasi kondisi-kondisi capaian dan keterampilan itu (lihat Skehan, 1988). Hal itu adalah komponen bahasa khususnya konteks. Corak yang sesuai akan menandai adanya derajat tingkat test yang mencerminkan atribut aktivitas dalam kehidupan riil. Kecuali jika langkah-langkah diambil untuk mengidentifikasi dan menyertakan. Corak seperti itu akan nampak kurang dalam membentuk pernyataan tentang suatu kemampuan yang memfungsikan kondisi-kondisi normal dalam target situasi masa depannya.

Dalam test literatur ada suatu penekanan kuat pada pentingnya tujuan test, dan tak satupun solusi dapat mengakomodasi skenario test ketereampilan membedakan area yang perlu dibuat untuk mengevaluasi kelompok teste yang berbeda dengan situasi target berbeda. Untuk mengukur cakupan bahasa cukup pada situasi: bagaimana, dengan siapa, dan mengapa bahasa digunakan, dan pada topik apa, dan dengan efek apa. Fakta bahwa pencapaian komunikatif diterima bersama dengan implikasi dari ketegasan test (lihat Alderson dan Urquhart, 1985b).

Peran konteks yang penting adalah sebagai penentu ttg kemampuan bahasa komunikatif dan pendekatan integratif terhadap penilaian sebagai lawan dari pendekatan decontextual. Bahasa tidak memiliki arti penuh jika adalah tanpa konteks (ilmu bahasa, discoursal dan sociocultural). Oller (1973, 1979) berpendapat tingkatan bahasa yang lebih tinggi adalah persepsi bahasa yang contextualized. Variabilitas pencapaian sesuai dengan jenis tugas yang dilibatkan, diimplikasikan terhadap panjangnya test dan jenis teks dan format yang cukup pada test (lihat Douglas dan Selinker, 1985; Skehan, 1987).

Keaslian tugas dan keaslian teks dalam test dihargai untuk menutupi permasalahan baik dalam definisi dan dalam perwujudannya. Jika tugas inauthentic adalah tercakup test ttg kemampuan bahasa komunikatif ada suatu

bahaya riil bahwa metoda yang diberikan bisa bertentangan dengan pengukuran perencanaan kita. Kita bisa mengakhiri mengukur kemampuan untuk mengatasi metoda itu bukannya kemampuan membaca, mendengarkan, menulis, berbicara atau berhadapan dengan suatu kombinasi ttg keterampilan ini konteks yang ditetapkan. Semakin asli tugas semakin sedikit kita perlu memperhatikan ini, sebagai contoh, cloze atau pilihan banyak. Test ttg kemampuan bahasa komunikatif harus langsung (mencoba untuk mencerminkan situasi “hidup yang riil”) dan tugas harus dilaksanakan perlu melibatkan proses ceramah realistik.

Bahasa tidak disederhanakan, “asli” teks harus digunakan sebagai masukan (lihat Widdowson, 1983) dan acuan cukup fungsional dan mempunyai petunjuk ttg bahasa. Sebagai tambahan perlu diperhatikan dimensi tugas seperti ukura teks dan proses waktu riil.

Netto berpendapat bahwa test yang berbeda perlu dirancang untuk tujuan dan instrumen yang berbeda, metoda atau isi tidak seragam.

Dalam menaksirkan kemampuan untuk saling berhubungan dengan lisan kita perlu mencoba untuk mencerminkan kealamiahannya interaktif percakapan normal dan mencoba untuk memastikan bahwa hal timbul balik dipertimbangkan untuk tugas test yang dimasukkan. Tugas harus diberikan dibawah waktu normal dan unsur tidak kemungkinan meramalkan interaksi lisan harus dikenali, untuk komunikasi asli boleh merujuk peserta lain. Calon boleh juga diberi tugas tertentu seperti tugas kelompok untuk mempertunjukkan suatu kemampuan mengatur interaksi dan untuk merundingkan maksud dengan teman bicara. Singkatnya apa yang diketahui dari teori ttg interaksi percakapan harus dirancang ke dalam tugas untuk menguji (lihat Bygate, 1987 dan Bandungan dan Bygate, 1990).

Legitimasi ttg test keterampilan terpisah ditanyakan, dan tentu saja uji coba keterampilan melalui story-line prosedur terintegrasi (lihat Slow 1936) sedang memperoleh kebaikan. Keaslian yang diragukan masih dapat dilihat mempunyai suatu peran (pelajaran) pelengkap untuk menambah keandalan test dan penaksiran ilmu bahasa secara rinci, tetapi titik sentral diberikan lebih mengarah pada usaha ke operationalise yang terintegrasi uji coba kemampuan bahasa komunikatif.

Arahkan test yang memerlukan suatu pencapaian teintegrasi dari peserta yang menyertakan komunikasi dalam bahasa yang realistis, situasional, budaya dan cenderung terbatas. Calon harus melaksanakan kedua-duanya, mudah menerima gagasan baru dan secara produktif dalam konteks relevan. Fokus adalah pada ungkapan dan pemahaman ttg maksud fungsional sebagai laawan suatu penguasaan bentuk yang terbatas. Gerak untuk mengarahkan pengujian lebih lanjut didukung oleh suatu perhatian, antara bahsa penguji tentang efek format permasalahan.

Efek format yang berhubungan dengan kemungkinan hasil percobaan mungkin berbeda dengan formrmat test yang dikerjakan, yaitu isi bacaan. Pengaruh metoda test mungkin terus meningkat jika belum dipahami sepenuhnya.

Dalam rangka menimbulkan pencapaian terbaik siswa adalah penting untuk memperkecil efek yang merugikan teknik pengukuran pada pencapaian. Jenis pencapaian yang ditimbulkan oleh metoda penilaian tertentu mungkin berbeda dari bahasa kehidupan riil yang digunakan.

Dalam area tanda-tanda, penilaian kualitatif dan holistic tt keterampilan dan usaha mengimplikasikan keandalan tes, perlu disampaikan. Permintaan suatu pendekatan criterion-referenced terhadap test kemampuan bahasa komunikatif dan penetapan score menuntut perhatian (lihat Bindley, 1986; Cziko, 1981; Hauptman et al., 1985). Langkah ahir menguji proses hasil percobaan harus ditunjukan ketika kita tinggal dugaan ttg kecakapan umum.

Kemampuan ilmu bahasa harus mrupakan bagian dari kemampuan komunikatif, cara yang ditempuh dengan menghubungkan satu sama lain, atau tentu saja bagaimana menghubungkan dengan kemampuan komunikatif, tetapkan dengan riset empiris. Sangat banyak pekerjaan perlu untuk dilaksanakan untuk membandingakn hasil perolehan dari pencapaian pada test kemampuan komunikatif didasarkan linguistik melalui kejadian pencapaian, sebelum seseorang dapat membuat statemen positif tentang yang terdahulu untuk dijadikan indikasi dalam situasi kehidupan riil. Tidak ada perbandingan realistis mungkin dapat dipercaya, metoda diselidiki untuk menilai kecakapan dalam melakukan tugas komunikatif yang relevan.

Bagi penguji paradigma komunikatif menjadi tekanan lebih besar untuk mengesahkan test oleh karena suatu keinginan untuk membuat test secara langsung yaitu tugas dan ukuran-ukuran. Klaim terhadap test untuk mengukur atau meramalkan pencapaian bahasa hidup riil harus bersifat sementara sampai kebenaran ukuran yang digunakan diperkuat. Ada tekanan yang harus ditetapkan sebagai kebenaran ukuran empiris dan teoritis yang terkandung di dalam paradigma.

Komitmen untuk membuat test yang komunikatif adalah memerlukan suatu derajat tinggi ketegasan pada tahap disain test dimana yang satu mempunyai kaitan dengan hasil yang diperlukan dan dilangkah evaluasi dimana yang satu sedang ditafsirkan hasil yang diperoleh (lihat Hawkey, 1982). Test komunikatif akan melihat secara radikal perbedaan dari beberapa test yang ada; tetapi mungkin ada pertimbangan kuat pragmatis untuk berusaha mempertunjukkan perbedaan dalam isi test, menandai rencana yang diterapkan dan hasil cara yang dilaporkan.

Test pada pengajaran berstatus ketidak pastian yang mendahului perlu dipertimbangkan serius jika test komunikatif mempunyai keuntungan terhadap pengembangan ttg kapasitas komunikatif dalam kelas (lihat Swain, 1985 Hughes, 1989) kemudian kita dapat lebih sedikit cemas akan kekurangan empiris atau teoritis ttg pengetahuan bahasa kita gunakan. Dengan cara yang sama kita jika kita dapat mencakup test yang dianggap sebagai praktek terbaik dan paling sesuai di kelas bahasa kenyataannya adalah bahwa banyak ditingkatkan. Prosedur yang diadopsi oleh dalam perancangan dalam keterampilan komunikatif dalam bahasa Inggris adalah penting.

1.4.4 Sebuah Paradigma Tes

Dalam test bahasa Inggris yang terbaru sebagai Bahasa Asing (Toefl) mengarah kemampuan/komunikatif yang menguji, cara test TOEFL yang dibuat lebih komunikatif tanpa dengan serius merusak hadirnya atribut psychometric (lihat Satansfield, 1986). Banchman (1986) yang menyelidiki ketiadaan ontoks yang berhubungan meteri TOEFL dan menyimpulkan bahwa: “mayoritas tugas

mengukur hanya kemampuan bersifat tatabahasa...dengan hanya suatu pencabangan illocutionary atau kemampuan sociolinguistik”.

Douglas (1986) dari interlanguage studi menerangkan variabilitas pencapaian yang disebutkan oleh prosedur penimbangan tugas dan oleh konteks dimana tugas dilaksanakan (lihat Douglas dan Selinker, 1985; Selinker dan Douglas, 1985 dan Skehan, 1984, 1987). Douglas berargumen bahwa jika TOEFL dan Test Percakapan Bahasa Inggris percakapan (Tse) telah ditinjau kembali arah tugas domain-specific mereka akan cocok kedalam suatu kerangka ttg kemampuan omunikatif. Pertimbangan juga diberikan terhadap keaslian bahasa yang digunakan(lihat Bachman 1986; Douglas, 1989) dan telah disetujui bahwa TOEFL tidak cukup memperhatikan dalam bagian mendengarkan yang normal ttg ceramah percakapan secara spontan, sebagai conto, keraguan, maupun corak interaksi yang normal seperti negosiasi ttg maksud bersama.

Dalam bidang pendidikan jasa: Layanan Ujian (ETS), princeton merupaka suatu organisasi yan sampai sekarang telah beroperasi dengan kuat dalam psychometric-strukturalis tradisional yang kini membuat testnya lebih komunikatif.

Dalam kelas bahasa, mayoritas test secara komersial yang tersedia sebagian besar masih structural-based (lihat Nolan-Woods, 1976, dan Coe, 1978; dan Allen, 1982). Kebanyakan bahsa mengajarkan coursebooks dan guru manual, apakah mereka memberikan nasihat pada test, pada umumnya menawarkan penyamarataan teoritissecara samar-samar dari kebutuhan praktis guru yang harus merancang test prestasi untuk digunakan di kelas. Nasihat yang spesifik dan yang ketinggalan jaman kadang-kadang disajikan pada berbagai poin terpisah, pendekatan atomistic tidak komunikatif, yang tidak cukup terhadap pengertian yang mendalam yang diperoleh melalui test riset dalam dua dekade terakhir.

Bantuan secara norma sangat kecil dsiapkan dalm bentuk test tugas untuk menguji tujuan atau dalam memilih format untuk menguji secara komunikatif. Hampir ada suatu kesenjangan pada bagaimana cara menginterpretasikan hasil percobaan ketika data telah dihasilkan.

Ada suatu kebutuhan mendesak bagi penerbit ELT untuk memperhatikan perkembangan bidang pengujian dan pengajaran bahasa komunikatif jika prakarsa baru tidak diciptakan. Riset Bidang Ilmu banyak penawaran hasil penyelidikan kecil informal oleh para guru, di berbagai konteks kelas (lihat Brindley, 1989).

Ada argumentasi kuat untuk mencoba UCLES/RSA. Dalam rancangan mereka dalam Keterampilan Komunikatif bahasa Inggris, test menggambarkan situasi praktek dalam kelas.

Satu-satunya perbedaan antara pengajaran dan pengujian didalam paradigma komunikatif berhubungan dengan jumlah bantuan yang tersedia di siswa dari gurunya atau panutannya. Bantuan yang secara normal tersedia dalam situasi belajar mengajar situasi, sebagai contoh, prompt, perumusan kembali pertanyaan, dorongan, koreksi dan kesempatan untuk mencoba lagi, dipindahkan menjadi test untuk pertimbangan keandalan pengukuran. Dalam hal ini test boleh jadi dipandang sebagai suatu langkah intermediate kelas dan target masa depan dimana calon akan harus beroperasi tanpa bantuan.

1.4.5 Cakupan yang Dijanjikan?

Pendekatan komunikatif bagi menguji bahasa telah diuraikan dan, secara umum, suatu pandangan positif tentang sejumlah permasalahan terkemuka didopsi menjadi suatu pendekatan yang perlu untuk ditunjukkan. Dalam rangka mencoba untuk merapat dengan permasalahan dalam hubungan dengan isu generalisasi hasil percobaan.

1.5 Permasalahan dalam Penghitungan

Selain dari permasalahan keandalan penanda serius, berhubungan dengan penilaian pencapaian (lihat bagian 4.3), isu yang utama mempengaruhi suatu adopsi “komunikatif” yang mendekati ke pengujian bahasa adalah generalisabilitas hasil yang diproduksi oleh suatu perjanjian.

Test dapat dilihat sebagai sampling instrumen yang menyediakan bukti untuk dasar kesimpulan luas di luar data tersedia. Bukti disajikan dengan test pencapaian harus relevan terhadap keseluruhan minat, test harus sah; harus dtabil

untukl dibuat sekitar suatui capaian calon dibagian daerah manapun, dengan kata lain, test harus dapat dipercaya.

Test komunikatif menyiratkan spesifikasi tugas pencapaian yang berhubungan erat kepada aktivitas pelajar secara praktis, konteks yang komunikatif dimana ia akan temukan sendiri. Permasalahn dalam generalisabilitas tugas mungkin memikirkan suatu bentuk test kecakapan untuk mengukur kemampuan mengkomunikasikan bahasa yaitu tergantung pada yang utama “sasran hasil test untuk diukur”.

Test performance-based memunculkan suatu target situasi permintaan spesifik dan sistematis adalah bahwa dipelukan tugas komunikasi tertentu, ketika kita tidak menggunakan bahasa yang sama untuk semua komunikasi. Dalam tes kecakapan, tugas tidak bisa dipisahkan dari sifat alami situau komunikasi dimana calon ditaksir.

Perhatian kita menunggu sampai keterangan empiris tersedia sebelum pembuatan. Identifikasi tugas ini hanyalah setelah pengujian jika mungkin untuk menetapkan sasaran hasil melalui riset empiris yang berdasar pada orang-orang riil yang mengatasi situasi riil, yang akan mengakui suatu contoh terpilih yang mewakili tugas operasional untuk menilai kemampuan capaian. Riset empiris diselenggarakan untuk menetapkan sasaran hasil, untuk mengidentifikasi tugas komunikatif da unsur dasar yang relevan yang mungkin keterampilan untuk suatu targrt populasi, permasalahan sampling, practicals, keandalan dan kebenaran. Faktor yang muncul dengan jelas adalah bahwa sangat meningkatnya ketegasan yang disempurnakan oleh keperluan analisis. Jenis ini setidaknya-tidaknya tidak mampu untuk menetapkan pentingnya variabel. Era (1978) dan Morrow (1979) menyarankan tujuan harus untuk membangunt ugas komunikasi ditirukan menyerupai kehidupan riil dan permintaan realistis dalam kaitan dengan pencapaian perilaku bahsa. Komunikasi tidaklah bersamaan dengan bahasa dan banyak komunikasi dalam ilmu bahasa. Sering kondisis-kondisi komunikasi nyata bukanlah replicable suatu situasi test, yang mana tidak dapat diidealkan.

Lebih lanjut, contoh tugas komunikatif harus menguasai isi dan kebenaran, tidak masih kekurangan generalisabilitas dalam kaitan dengan yang tugas

komunikatif lain. Apakah penilaian capaian pada tugas ini dibuat di bawah batasan sosial dan ilmu bahasa tertentu dan tidak relatable “kemampuan karakteristik?” dengan kata lain, jika suatu pemilihan dibuat, contoh diambil dari suatu daerah, dipastikan bahwa contoh adalah cukup?.

Kelly (1978, p.226) mengamati bahwa semua macam test adalah suatu latihan dalam sampling dan suatu usaha dibuat untuk menyimpulkan kemampuan siswa dalam hubungan dengan capaian mereka secara umum. Itu semua adalah hasil siswa yang diharapkan untuk mengetahui perilaku sebagai hasil siswa yang diharapkan untuk mengetahui perilaku sebagai hasil abahan pengajaran dalam suatu test prestasi. Keandalan suatu test dalam konsep ini adalah tingkat score pada test yang mengindikasikan kestabilan ttg kemampuan calon dalam hubungan dengan pengetahuan alam semesta yang lebih luas, pencapaian. . ia menunjukkan (p.230) bahwa banyaknya permasalahan komunikasi yang membedakan calon harus memecahkan kondisi-kondisi dunia nyata adalah sama besar seperti permutasi dan kombinasi yang diproduksi oleh nilai-nilai variabel dalam sort, jenis pesan, konteks situasi dan kondisi-kondisi capaian yang mungkin ditemui.

Atas dasar capaian pada item tertentu, seseorang hendaknya, menjadi berhati-hati dalam menggambarkan; menarik kesimpulan tentang suatu kemampuan untuk menangani tugas komunikasi serupa.

Morrow (1977, p.53) sadar akan permasalahan penghitungan. Ia dengan ringkas tapi jelas memperkenalkan masalah seperti ini:

Seluruh intisari suatu pendekatan komunikatif adalah untuk menetapkan situasi tertentu dengan corak konteks tertentu, dll.. dalam rangka menguji kemampuan calon untuk menggunakan bahasa yang sesuai dalam kaitan dengan spesifikasi tertentu. Ketika berharap agar prosedur hendak dibahas dalam terminologi itu, mereka tidak bisa mengungkapkan kemampuan calon untuk menghasilkan bahasa yang sesuai.

Alderson (Alderson dan Hauhes, 1981, p.59) juga menerima bahwa untuk mengikuti paradigma yang komunikatif, seseorang perlu menggambarkan apa yang para siswa harus lebih dulu lakukan atas bahasa dalam suatu rangkaian atau situasi-situasi spesifik.

Dalam rangka membuat ramalan pencapaian siswa yang berhubungan dengan alam semesta, tugas yang besar akan nampak diperlukan untuk merasakan capai, konflik dengan seketika dengan permintaan efisiensi test. Tugas lebih besar dan semakin realistis terhadap materi test, perhitungan bukanlah rumit: sekalipun kita tidak bisa menyamaratakan capaian dalam satu situasi menuju pencapaian diberbagai situasi, jika kita dapat mengatakan tentang sesuatu capaian dalam satu situasi, kita sudah mendapat kemajuan, dan jika kita dapat mengatakan sesuatu yang penting sekitar pencapaian dalam situasi target yang lebih baik, siswa harus melaksanakan samping bukti hubungan antara peramal dan meramalkan, atau hubungan yang theoretised antara kemampuan dan capaian.

Langkah yang mudah untuk mengidentifikasi “satu situasi” didasarkan pada ramalan kita. Contoh pengembangan dari EAP tentang membaca, untuk sampling dalam suatu test terjadi pemilihan teks yang dipertunjukkan keterampilan pengertiannya. Ada beberapa bukti literatur yang tidak menguntungkan untuk diuji pada pengertian teks (lihat Damm, 1983A; Alderson dan Urquhart, 1985A, 1985B). perbedaan implikasi test adalah jelas bisa diidentifikasi. Sehingga ada suatu kebutuhan mendesak untuk penyelidikan lebih lanjut kedalam pengujian bahasa untuk tujuan ditetapkan.

Morrow (1977) mengamati kasus ttg test bahasa konvensional yang mengarah pada pengukuran penguasaan kode bahasa, penghitungan akan menampakkan sedikit permasalahan. Sistem fonologi yang bersifat tatabahasa dapat dikendalikan dan dibatasi dan sumber daya yang berhubungan dengan kamus dapat dilimited. Bahasa terdiri dari satu jumlah unsur-unsur terbatas, dan test penguasaan tentang unsur-unsur ini sangat kuat dari segi pandangan yang bersifat prediksi. Davies (1978, p. 225) mengatakan bahwa argumentasi berpihak pada kemampuan test ilmu bahasa (terpisah dan integratif) adalah inti belajar bahasa... tatabahasa adalah berkaitan kuat lebih jauh dengan generalisabilitas dibanding corak bahasa lain.

Kelly (1978) berargumen bahwa suatu penguasaan ilmu bentuk kata kerja bahasa Inggris lengkap adalah terhadap keseluruhan sasaran untuk mampu

mengkomunikasikan bahasa Inggris. Kelly (1978, p.17) menguraikan bahwa “pengetahuan fungsional relatif penting diberbagai struktur dalam suatu bahasa”.

Dengan kekurangan kemampuan akannampak keliru untuk membuat klaim bahwa para siswa bisa merusak suatu bahasa atas dasar score pada test kosa kata atau sintaksis terpisah. “kemampuan untuk mengkomunikasikan bahasa” melibatkan lebih dari suatu manipulasi ttg pola syntactic tertentu dengan isi kamus tertentu. Karena itu, akan nampak bahwa ada suatu kebutuhan untuk mencoba memikirkan alat ukur yang dapat menilai kemampuan capaian.

Kelly (1978,p.239) mengusulkan dua langkahmendekati ukuran-ukuran kebenaran, keandalan dan efisiensi:

Langkah pertama melibatkan pengembangan suatu test langsung yang secara maksimal dapat dipercaya dan sah, dan kebenarannya tidak efisien. Langkah kedua meminta pengembangan ttg efisien, karenanya yang tidak langsung, test ttg kebenaran. Test kebenaran yang tidak langsung (diharapkan) untuk ditentukan dengan acuan ketugas langsung dengan jelas, dapat dipercaya dan sah test tidak efisien untuk membangun yang dimasalahkan.

Test kecakapan terkait dengan pemusatan perhatian pada bahasa yang digunakan individu, test kemampuan menangani aspek bahasa itu yakni sistem berkenaan dengan fonologi dan yang bersifat tatabahasa. At TEEP menyebutkan test tatabahasa tidak menawarkan apapun informasi tambahan penggunaan berdasarkan komponen.

Morrow (1978, p.152) berargumentasi bahwa suatu model untuk tugas komunikatif global boleh ditunjukkan, untuk itu tugas, harus dikerahkan untuk dilengkapi. Keterampilan kemampuan pencapaian adalah suatu analisa capaian dalam terminologi operasional, dan performance-based. Aplikasi mereka meluas dan kreativitas mencerminkan aspek yang dipahami oleh kemampuan. Dengan cara ini mereka menawarkan suatu pendekatan perhitungan.

Ia menyatakan bahwa (p.153): “analisa tugas menghasilkan suatu kemampuan keterampilan” dan membantah penilaian yang konsisten, oleh arena itu hasil data relevan ke seberang spektrum tugas global, dan pencapaian tunggal.

Morrow (1979, p.153) mengatakan bahwa solusi pengembangan test mengukur keseluruhan capaian dalam hubungan dengan suatu tugas yang ditetapkan dan strategi dan keterampilan yang telah digunakan dalam mencapai itu.

Dalam tes mendengarkan dan membaca pengertian, agak lebih sukar untuk diselesaikan... untuk menyimpulkan strategi dan keterampilan yang dipakai, dengan menyediakan basis untuk meyakinkan. Test keterampilan memiliki kelemahan untuk menangani individu yang memungkinkan keterampilan, namun masih tidak bisa komunikasikan secara efektif.

Masalah lain adalah bahwa tidak sama sekali mudah untuk mengidentifikasi keterampilan; tidak ada petunjuk untuk meaksir penyelesaian tugas komunikatif tertentu. Morrow berasumsi bahwa kita tidaklah hanya mampu menetapkan kemungkinan keterampilan, tetapi untuk menguraiakan hubungan yang ada antara bagian dan keseluruhan adalah suatu cara yang akurat 'terpisah' untuk tugas yang komunikatif. Ada rumus pemilikan dan kemampuan keterampilan $X+Y+Z=$ penyelesaian tugas komunikatif.

Jalan keluar yang pragmatis adalah menguji untuk mengejar suatu pendekatan pengesahan etnografi digambarkan dibagian 2.1.2. data bisa dikumpulkan untuk bantuan menentukan materi yang terbaik yang diperlukan (lihat Aslain, 1985; Cohen, 1985; dan Jones dan Friedl, 1986).

Nasihat bisa diambil dari para profesional untuk menafsirkan kepantasan materi test untuk populasi yang diharapkan (lihat Douglas dan Pettinari, 1983). Yang terbaru IELTS revisi proyek telah mengadopsi strategi bermanfaat ini.

Masalah penghitungan adalah ttg gejala replicable. Peneliti di daerah dapat secara langsung menghadapi apa yang mereka ingin selidiki, merumuskan hipotesis dan eksperimen pengulangan, memverifikasi atau memalsukan hipotesis mereka. Oleh karena permasalahan paradigma yang ilmiah sulit untuk pengukuran bidang pendidikan.

Hawkey (1982) menguraiakan paradigma ilmiah yang klasik sebagai hypothetico-deductive yaitu metodologi yang merumuskan hipotesis, menguji melalui pengamatan atas perilaku suatu sampel acak untuk populasi target, yang

dikuti dengan suatu analisis hasil statistik. Pendekatan ini tidak pantas untuk kecakapan yang menguji berbagai konteks dan pada berbagai tingkatan. Tanggungjawab perlu untuk diambil, sebagian tidak dapat diramalkan, semua saling berinteraksi konteks kultur sosial. Dengan begitu ada suatu masalah sampling tugas, suatu masalah kebenaran.

Tidak sama dengan paradigma ilmiah yang diuraikan oleh Hawkey yaitu jika populasi target para siswa adalah sementara, secara luar dibubarkan dan memvariasi dalam kaitan dengan keadaan dapat masuk, kekuatan sampling karena terpaksa harus oportunistik. Ini adalah suatu masalah sampling populasi, suatu masalah keandalan.

Hawkey (1982, p.16) menguraikan bahwa 'evaluasi bersifat memberi paradigma, uraian gejala kompleks, resolusi ttg corak penting, dan pengertian hubungan untuk membatasi kerangka deskriptif untuk menetapkan tugas komunikasi keterkaitan ke para siswa dalam suatu konteks yang ditetapkan, sebelum konstruksi test (lihat Bendungan, 1983B).

ESP adalah suatu cerminan ttg kekuatan bahasa Inggris untuk tujuan yang ditetapkan secara khusus atau spesifik. Menekankan pengajaran dan pengujian dan tidak pernah secara total.

Kekurangan target situation analisa dispesifikasikan ebutuhan tes bahasa (lihat Bendungan, 1983A). barangkali bahaya yang lebih besar adalah bahwa ada kecenderungan untuk analisis untuk mengakui suatu tidak sebanding jumlah waktunya dan sumber daya yang tersedia untuk riset; sering pada biaya pengembangan test.

Weir (1983A, 1983B), pengujian yang dihubungkan test bahasa Inggris untuk Tujuan Bidang pendidikan (Teep) secara penuh didokumentasikan (lihat juga catatan tambahan I). hasil analisis keperluan mempengaruhi adopsi format test tertentu (dikte, pengintegrasian aktivitas membaca, mendengarkan dan menulis) dan memperjelas cakupan keterampilan untuk diuji dan ukuran-ukuran penilaian untuk dipekerjakan. IELTS bersandar pada tenaga ahli 'berkomentar untuk menilai'keaslian dan menulis regu, tidaklah dibatasi oleh purbasangka spesifikasi (lihat catatan tambaha V).

Dalam pengembangan test dimasa datang keseimbangan perhatian harus dispesifikasi kedalam perwujudan test dan mensahihkan yang keterbelakangan untuk menetapkan penulisan item yang akan diukur. Penekanan harus pada pengembangan test dan pengesahan dibandingkan pada analisa kebutuhan untuk menciptakan spesifikasi test.

Langkah yang rumit dalam pengembangan test ketika spesifikasi diterjemahkan kedalam suatu perwujudan test. Test memperlihatkan kualitas kebenaran, efisiensi dan keandalan yang diuji. Kualitas ini perlu untuk ditentukan kedua-duanya menurut mutu berdasar purbasangka dan dengan pengalaman suatu posteriori.

BAB III

PERTIMBANGAN DALAM MENDESAIN TES BAHASA

3.1 Konsep Validitas

3.1.1 Membentuk Validitas

Konsep validitas (apakah menggrjakan 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 memasukan '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 bukanya 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 Englis-

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 kecakapan bahasa tidak hanya untuk pengujian komunikatif. Capaian bahasa komunikatif dengan jrlas suatu dimensi kecakapan bahasa. Tetapi kemampuan bahasa adalah suatu dimensi kecakapan 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 kecakapan 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.

BAB IV KONSTRUKSI TES BAHASA

Dalam bab ini diuraikan tentang tahap-tahap pengembangan suatu tes. Secara eksternal, para siswa diuji secara besar-besaran dipersiapkan untuk kelas-kelas bahasa yang akan memiliki satu efek penting dalam kehidupan mereka dan penting untuk mengevaluasi bagaimana kekakuan test-test ini telah dikembangkan dan seberapa jauh test-test itu bertemu dengan kriteria-kriteria ukuran keandalan, efisiensi dan kebenaran yang dibahas dalam bab dua.

Ada empat langkah-langkah yang kini diterima sebagai praktek terbaik dalam pengkonstruksian dan pengesah tes, yakni desain, pengembangan, operasionalisasi dan monitoring (lihat Courchene dan de Bagheera, 1985; Carroll dan Hall, 1985).

4.1 Desain Test

Di dalam bab satu kita menguji dengan teliti apa yang bisa dilibatkan dalam rancangan suatu test antara paradigma komunikatif dan ukuran tertentu ditandai bahwa test seperti itu diharapkan bisa untuk disajikan, contohnya haruslah interaktif; langsung secara alamidengan tugas mencerminkan aktifitas proses tulisan yang realistis; teks-teks dan tugas-tugas harus sesuai dengan keadaan populasi target diharapkan; artinya kemampuan harus dicontohkan antar pengertian dan pengembangan konteks dan pengujian harus didasarkan pada suatu yang tersirat spesifikasi yang prioritas. Di bab dua kita membahas bagaimana kevalidan pengujian perlu dirancang dengan mapan, reabilitas yang kuat dari pilihan test yang terpilih dan prosedur yang ditetapkan.

Untuk mengejar paradigma komunikatif, tugas-tugas semestinya, seawal mungkin, dimasukkan ke dalam pelaksanaan ujian langsung dengan persyaratan yang dengan teliti dan menguraikan aspek penting dari capaian kegiatan dan dilaksanakan dengan keadaan n secara normal.

Awal pengujian ini adalah gambaran umum parameter secara spesifik dari test yang ingin dicapai, keadaan populasi, terlepas dari pendapat-pendapat yang berbeda pada metodologi yang sesuai untuk dilakukan. Parameter-parameter ini berguna bagi para penguji sebagai suatu daftar pemeriksaan yang mencocokkan kinerja mendasarkan tugas ujian yang dapat dievaluasi. Jika keinginan itu untuk mensimulasikan dalam situasi ujian para siswa beraktivitas yaitu dihadapan dengan dunia nyata dengan demikian penting untuk memiliki suatu dasar yang sistematis untuk gambarkannya. Jika suatu gambaran umum parameter dapat digunakan untuk proses situasi yang diharapkan diwujudkannya yang kemudian bisa digunakan untuk mengevaluasi derajat persamaan antara tugas-tugas pengetesan dan pilihan aktivitas yang dilibatkan, atau nampak untuk dilibatkan, dalam keadaan yang sebenarnya.

Parameter-parameter itu penting untuk mengumpulkan informasi termasuk:

- a. Aktivitas - bagian tugas para siswa harus mengatasi sewaktu ambil bagian dalam kejadian.
- b. Setting - secara fisik dan konteks psycho-social dari kegiatan-kegiatan itu.
- c. Interaksi - calon perangkat peranan dan calon hubungan sosial dilibatkan.
- d. Instrumen - medium, gaya dan saluran aktivitas dalam kejadian.
- e. Dialek- calon berbagai dialek dan aksen-aksen diperlihatkan.
- f. Keterampilan yang berpeluang - keterampilan dasar yang muncul penting yang berpeluang bagi siswa untuk dilaksanakan dalam berbagai aktivitas.

Data empiris bisa dihasilkan pada jumlah dan kesukaran yang ditemukan dari kinerja dalam aktivitas yang relevan dalam keadaan yang diharapkan di bawah batasan-batasan yang berlaku. Yang lebih berat, langkah-langkah harus diambil untuk menetapkan hubungan yang penting pada kegiatan itu ke kegiatan yang memberikan kepuasan dalam keadaan yang diharapkan.

Dalam pendekatan komunikatif pada pengujian bahasa seyogianya menjadi satu penekanan bukan pada ketelitian ilmu bahasa, tetapi pada

kemampuan fungsi secara efektif melalui bahasa dalam pengaturan tertentu dan konteks-konteks tertentu. Hal ini melibatkan dugaan bahwa aktivitas linguistik dalam ujian mesti berbagai hal dan di bawah keadaan yang mendekati pada hidup yang sebenarnya.

Davies (1978) berpendapat bahwa kita perlu melakukan peningkatan bahwa konteks pengujian yang kita miliki berdasarkan tatabahasa, di dalam pengertian bagaimanapun pembuatan materi test lebih realistis. Rea (1978) berargumentasi bahwa dalam hal mendesain test melibatkan hal yang bersifat komunikatif secara langsung yang akan memecahkan dalam kehidupan nyata dan menanyakan hal yang langsung realistis erkaitan dengan mereka dengan menggunakan istilah perilaku-perilaku berbahasa.

Isu nampak apakah ada dimensi-dimensi penggunaan bahasa yang buka bagian dari test-test dan suatu perspektif omunikatif perlu dipersatukan, karena penting untuk diperlihatkan kepada mereka. Sementara untuk menilai komunikasi hidup riil tidak bisa dijawab di dalam situasi test, kita masih perlu mencoba dan membuat test yang sama yang mungkin realistis seperti dengan kaitannya situasi itu.

Hanya jika kita mencoba membuat test yang mendekati mungkin dengan tugas para siswa yang dihadapi dalam target dan kondisi yang dilaksanakan secara normal, kita sanggup untuk menilai apakah lebih sedikit ukuran langsung kemampuan yang sama dapat melengkapi informasi yang serupa tentang kinerja siswa. Ada suatu kebutuhan test sama yang mungkin seperti sebelumnya untuk membandingkan keefektivitasan dan test integratif yang untuk mengukur hal yang sama. Dalam test-test yang komunikatif, kita perlu memberikan kesempatan. Widdowsom (1978; p.80) menyebut 'asli' penggunaan bahasa, yaitu, meletakan pelajar di dalam posisi-posisi diman 'ia diwajibkan untk berhubungan dengan'... kejadian penggunaan bahasa 'dengan cara 'aktivitas komunikatof yang normal'. Jika penguji merasa terkait dengan kondisi komunikasi seperti riil yang mungkin dal testnya, Hawkey membantah (1982, p.164) bahwa mereka perlu menguraikan apa yang terjadi.

Karakteristik di bawah bersifat ilustratif dibanding yang sebenarnya. Bagaimanapun, ketidak mungkinan teori yang cukup dari bahasa digunakan, dipaksa untuk memperhatikan karakteristik itu dikenal sebagai bagian penting bagi praktisi-praktisi dalam pengajaran bahasa.

Tidak ada klaim yang komprehensif maupun mendeskreditkan dari kategori-kategori yang ada didalamnya. Sebagai tambahan karakteristik itu lebih yang sesuai kepada perantaranya dibanding yang lain. Yang jelas semakin banyak suatu test mereflesikan karakteristik komunikatif dinamis yang sesuai target aktivitas, semakin banyak relevan perilaku bahasa yang muncul. Meskipun tidaklah mungkin, untul pertimbangan praktis, menyertakan semua fitur ini kedalam suatu test, acuan karakteristik ini akan menyediakan suatu ukuran yang mana ujian boleh jadi dinilai dan dibandingkan dengan test-test yang dirancang untuk menimbulkan perilaku kinerja yang sama.

Daftar di bawah karakteristik komunikatif yang perlu dipertimbangkan dalam merancang test komunikatif. Semuanya berasal dari pertanyaan yang diatur untuk guru sekolah bahas (lihat Weir, 1983a), dan Roger Hawkey (1982) dan eith Morrow (1977, 1979). Karakteristik yang bisa diharapkan adalah:

- a. Konteks realistis - pertanyaan-pertanyaan ujian harus dihargai dengan keadaan peserta.
- b. Relevansi kesenjangan informasi - peserta haru memproses informasi baru sebagai kekuatan dalam situasi hidup yang nyata.
- c. Intersubjectivitas - pertanyaan-petanyaan perlu melibatkan peserta baik sebagai penerima bahasa maupun penutur bahasa. Sebagai tambahan bahasa yang dihasilkan oleh peserta harus dimodifikasi sesuai dengan harapannya.
- d. Lingkup pengembangan aktivitas bagi para peserta- pertanyaan-pertanyaan memberi kebebasan dan komunikatif juga tidak memungkinkan bentuk pertanyaan yang menebak-nebak.
- e. Kesempatan memonitoring sendiri bagi para peserta - pertanyaan-pertanyaan memberikan kesempatan pada peserta menggunakan

strategi untuk mengevaluasi efektifitas dan berkomunikasi yang dengan suatu peristiwa.

- f. Proses masukan ukuran yang wajar - ukuran dan lingkup aktivitas mesti sama dengan proses masukan secara normal yang diharapkan mereka.
- g. Waktu normal membatasi pengerjaan - pertanyaan harus dicapai di bawah batasan-batasan waktu normal.

Hawkey (1982) membuat desain pertanyaan lanjutan dalam kumpulan penskoran untuk menetapkan bentuk-bentuk kejadian tertentu. Menetapkan pertanyaan test suatu uraian dalam bentuk ilmu bahasa secara objektif, gaya penulisan dengan target bentuk pertanyaan terkait mendekati pada bentuk aktivitas berbahasa pada sasaran yang sama, yang kedua itu, memungkinkan penguji merencanakan pengukuran penilaian kinerjanya terhadap bentuk-bentuk yang tidak bisa dipisahkan dalam tugasnya sendiri. Kekuatan bentuk pertanyaan test berguna dengan mempertimbangkan:

- a) ukuran test - panjang teks, receptif atau prodektif, dilibatkan;
- b) kompleksitas tatabahasa dan cakupan alat-alat kohesi diperlukan;
- c) cakupan fungsional - tingkat variasi ilokusioner dilibatkan; serta
- d) cakupan petunjuk - luas kedalaman pengetahuan dihubungkan dengan klamus yang diperlukan untuk menangani aktivitas.

Dalam konstruksi test yang komunikatif sampai saat ini, ditetapkan parameter deskriptif secara umum, populasi yang target harus dilaksanakan menekankan pada kebanyakan yang sudah diterima. Mengumpulkan informasi karakteristik komunikatif yang dinamis diberikan lebih meragukan pada teori komunikasi dan sungguh praktis dan berbagai kesulitan teoritis dalam menetapkan bentuk pertanyaan selain dari ukuran.

Penetapan suatu kerangka dari kategori uraian test yang komunikatif dapat ditemukan pada Hawkey (1982) dan Weir (1983b). suatu usaha membuat spesifikasi perancangan suatu test di dalam bahasa Inggris untuk Educational Purposes dan perwujudannya didokumentasikan di Weir (1983a).

Pembuatan yang ketat dalam mendukung bagian yang tidak realistis dalam berbagai situasi yang sudah berjalan. Mungkin sulit untuk menyiapkan isi spesifikasi, luas dan kedalamannya akan sungguh-sungguh bergantung pada hal-hal praktis seperti waktu yang tersedia, jumlah pelaksanaan dan ahli yang ada. Kebutuhan analisa dibatasi pada pertanyaan pendek atau suatu perwakilan yang menggambarkan situasi yang dicapai untuk membuka peluang pada situasi keterampilan diperlukan.

Pada bagian tertentu yang baik diteliti, pengalaman dan penilaian para ahli boleh jadi satu sumber yang valid dengan data yang sama untuk menganalisis kebutuhan dan mempersiapkan suatu test yang khusus. Bahayanya dari jenis ini bagaimanapun penganalisan tidak diabaikan.

4.2 Pengembangan Test

Apabila spesifikasi memuaskan sudah dipersiapkan harus dikembangkan dalam bentuk test dengan menyertakan teks (pertanyaan yang jelas dan teks yang harus dipecahkan dalam tahap disain). Bentuk harus terpilih dan soal ditulis mencerminkan sesuatu yang mungkin aktivitas dari situasi yang diharapkan.

Pilot test perlu pengesahan lebih lanjut periksa pada tahap ini dengan mengundang para ahli dalam bidang bahasa untuk menafsirkan pantas tidaknya teks, bentuk dan materinya. Setiap perkembangan perlu kejelasan pembuatan dari kecermatan penelitian. Metoda pengujian pengadministrasi dan pembuat test perlu diformalisasikan dan diujicobakan sebelum dilaksanakan.

Uji kepedulian memastikan bahwa contoh peserta-peserta dalam percobaan itu cukup karena memerlukan derajat ketepatan seperti contoh populasi target yang mungkin. Pendapat juga diperlukan guna penentuan test, dari para ahli mengenai isi: kepantasan, tingkatan, dll. (lihat Weir 1983a, p.346).

Keterangan empiris dari kevalidan eksternal ujian perlu dimapankan melalui langkah-langkah pengesahan yang bersamaan (lihat Section 21). Reliabilitas perlu untuk diperkirakan seperti yang digambarkan diatas (lihat Section 22). Analisis pada soal, tugas dan membangun tingkatan perlu dilaksanakan untuk memastikan bahwa materi dan soal-soal bersifat memuaskan

dijawab. Pengujian perlu ditinjau kembali dari sudut kualitatif dan kuantitatif untuk menghasilkan reliabilitas, validitas dan efisiensi.

4.3 Pelaksanaan

Tahap yang kaitannya dengan bersiap-siap menghadapi dan melaksanakan. Pertama, melibatkan persiapan, pencetak dan pendistribusian informasi yang sesuai bagi peserta, lambang/institusi pelaksana, pengawas ujian dan pengadministrasi.

Keputusan diambil terlebih dahulu sebagaimana peserta-peserta dikenalkan pada bentuk-bentuk test, terutama jika test yang baru; menunjukkan suatu pengawalan yang sulit bagi test yang sebelumnya. Contoh pengujian bisa diberikan untuk memungkinkan peserta mengenali dirinya terlebih dahulu dengan langkah-langkah pengujian.

Contoh-contoh bentuk test dan penjelasan bagaimana menjawab harus dipahami (dalam bentuk pedoman-pedoman). Jika ini tidak disediakan terlebih dahulu perlu memberi latihan pada peserta pada hari yang ditentukan.

4.4 Monitoring

Hasil-hasil dari tiap lembar ujian secara hati-hati dimonitor dan idealnya disimpan dalam data base. Pertimbangan perlu dilakukan pada revisi-revisi masa depan dan berkelanjutan.

Perlu memonitor pengaktualisasian dari test dalam pengajaran di dalam kelas-kelas sebelum. Kajian yang bersifat prediksi harus sesuai guna melihat apakah suatu kecakapan test mengidentifikasi mampu dan tidak mampu. Dampak dari suatu test yang baru perlu juga dimonitor seperti juga resepsi ujian oleh penerima lembaga menggunakan informasi dari hasil percobaan.

Sesungguhnya tidak semua langkah-langkah ini tercakup dalam praktek kelas normal. Semuanya berfungsi untuk menggambarkan, bagaimanapun, kompleks dan proses yang panjang yang perlu diamati untuk menjamin bahwa menguji dibuat valid, efisien dan dapat dipercaya. Guru kelas dengan tidak biasanya mempunyai waktu atau kecakapan untuk menyesuaikan diri pada

langkah-langkah yang diuraikan diatas dan apakah dimungkinkan untuk mencapai manfaat dalam batasan yang dikenakan oleh sekolah tersebut.

Jawaban yang sederhana adalah karena paling sedikit dimungkinkan untuk lebih ketat didalam desain tugas test dan jika yang dilaksanakan dalam konstruksi awal penting apakah langkah pengembangan test, banyak dapat dicapai. Sepanjang langkah guru kelas nampaknya sanggup melakukan kita berkonsentrasi pada bab empat, kita mengetahui alternaif-alternatif utama untuk diuji coba keterampilan bahasa dan menguraikan secara singkat keuntungan-keuntungan dan kerugian-kerugian dari mengadopsi bentuk-bentuk test tertentu.

BAB V

METODE TES BAHASA

Metode-metode digunakan untuk membangun test tapi bukan pada test itu sendiri. Meskipun mungkin untuk membicarakan tidaknya satu test tapi nyatanya tidak bagi metode. Prosedur pilihan ganda mungkin menghasilkan test yang valid vada suatu realisasi tapi tidak padarealisasi lain. Ini merupakan kasus untuk semua metode dan akan mengingatkan kembali pembahasan selanjutnya terhadap keuntungan dan kerugian potensial dari metode-metode yang berbeda berikut ini.

Pendekatan-pendekatan yang berbeda terhadap ujian bahasa adalah diuraikan di dalam bab I, referensi dibuat bagi efek yang memungkinkan dari metode test terhadap skor test. Terdapat bukti yang sama dalam literatur (lih. Murphy, 1978, 1988; Porter, 1983; Weir, 19...; Boniakowskan, 1986; Alderson dan Urganast, 1985a) bahwa format test mungkin mempengaruhi ferformansi siswa, membatasi tingkat pengetahuan mengenai efek format test terhadap satu-satunya pendekatan praktis saat ini yaitu untuk melindungi kemungkinan efek format yang mengmbangkan dasar test dengan lebih luas melalui penggunaan variasi format yang valid, praktis, dan reliable bagi ujian tiap keterampilan.kita belum membahas beberapa metode test yang berbeda secara detail pada penggunaannya. Oleh karena itu bab ini memberikan laporan singkat tentang jenis-jenis utama format test dan beberapaa *Manfaat* dan *Kerugian* potensialnya. Ini dimaksudkan untuk memberikan referensi pada pedoman untuk kontruksi test yang akan datang.

Sebagaimana kaidah umum, ia adalah yang terbaaik untuk menilai dengan menggunakan fariatif format tersebut, skor diambil adalah skor gabungan untuk melaporkan tujuan. Syarat atau ketentuan utama untuk ujian menggunakan kerangka komunikaatif adalah bahwa tugas test akan menunjukkan pengolahan

percakapan nyata sejauh mungkin dan mencakup susunan keterampilan-keterampilan yang mungkin yang telah diperkenalkan (lih. Appendix I sebagai contoh dari pendekatan ini dalam test TEEP). Penting bahwa test-test yang dikembangkan dalam paradigma ini akan mempunyai efek washback yang kuat terhadap praktek di kelas bahasa.

5.1 Pendahuluan

5.1.1 Pertanyaan-pertanyaan Pilihan Ganda (MCQS)

Saran untuk penyusun soal-soal pilihan ganda juga dapat dipakai pada penyusunan test-test pemahaman menyimak, struktur, dan osa kata. Ini semua akan dibahas dalam bab ini.

Soal ujian pilihan ganda biasanya diberikan dengan cara di mana kandidat harus memilih jawaban dari pilihan-pilihan yang diberikan, hanya satu yang benar. Proses penilaian semuanya objektif karena penilai tidak diperbolehkan untuk membuat pertimbangan ketika menilai jawaban kandidat, mufakat telah dicapai seperti terhadap jawaban yang benar dari tiap soal. Pemilihan dan penyusunan soal merupakan proses yang subjektif dan keputusan tentang mana jawaban yang benar merupakan persoalan pertimbangan yang subjektif pada penulis.

Manfaat

1. Dalam test pilihan ganda terdapat reliabilitas penialia yang hampir sempurna. Nilai-nilai yang ada pada format subjektif, tak bisa dipengaruhi oleh pertimbangan personal atau oleh keistimewaan-keistimewaan penilai. Selain dapat dipercaya, penilaiannya itu sederhana, lebih cepat dan sering lebih efektif daripada bentuk lain dari test tertulis.
2. Karena soal itu bisa diuji coba dengan hampir mudah, biasanya memungkinkan untuk memperkirakan dulu tingkat kesulitan dari tiap soal. Free-test juga memberikan informasi tentang tingkat yang tiap soal tambah secara positif terhadap apa yang test pada umumnya nilai. Dwiarti dalam susunan kata-kata dari soal-soal mungkin juga diungkapkan dalam analisis data free-test dan bisa dijelaskan atau dijernihkan dalam tests yang sebenarnya.

3. Format dari soal test pilihan ganda yaitu seperti tujuan-tujuan dari penyusun test itu jelas dan tegas; para kandidat mengetahui apa yang mereka perlakukan. Pada format-format open-ended dwi arti dalam susunan kata-kata pada pertanyaan mungkin kadang-kadang menuntun para kandidat yang mengatur waktu tambahan menjawab pertanyaan berbeda dengan apa yang diharapkan penguji.
4. Pada format open-ended, contohnya pertanyaan-pertanyaa jawaban singkat, calon atau pesertaharus menyebutkan keterampilan menulis. Tingkat yang mempengaruhi pengukuran-pengukuran yang akurat terhadap ciri bawaan dinilai bukan ditentukan. Test-test pilihan ganda menghindari kesulitan khusus ini.

Kerugian

1. Bagamanapun juga terdapa bebrapa masalah yang brhubungan dengan penggunaan format ini. Jika seorang kandidat menjawab soal pilihan ganda karena bebrapa kekurangan dalam pertanyan, lembar jawaban tempat ia menjawab jawabannya tidak akan mengungkapkan kebenaran ini. Dan lagi, kita tidak tahu apakah kegagalan kandidat seharusnya dari kurangnya pemahaman terhadap teks atau terhadap pertanyaan. Seorang kandidat mungkin menjawab beanar sebuah soal dengan menyingkirkan jawaban-jawaban yang salah, suatu keterampilan berbeda dari kemampuan memilih jawaban yang benar pada tempat pertama.
2. Skor-skor yang dicapai dalam test pilihan ganda, seperti dalam test benar-salah, mungkin dicurigai karena calon telah menerka semua atau bebrapa jawaban. Ini berakibat penyempitan skor. Format dari test-test ini mendorong kandidat untuk menebak semuanya, kadang-kadang perlu dipertimbangkan untuk mengambil langkah mengecilkan hati kandidat untuk melakukan itu. Mungkin juga untuk menyelesaikan soal tanpa melihat teks yang ada, jika ini terjadi, apapun yang diujikan tidak bisa menjadi pemahaman teks.
3. Test-test pilihan ganda lebih lama dan lebih mahal dan lebih sulit untuk mempersiapkannya daripada beberapa ujian open-ended seperti komposisi.

Banyak soal yang harus ditulis dengan hati-hati oleh penulis soal yang terlatih dan meskipun telah diuji coba sebelum digunakan pada ujian format.

Tiap soal harus di edit dengan teliti untuk memastikan bahwa:

- Tidak ada informasi yang berguna.
 - Ejaan, grammer (tatabahasa) dan tanda bacanya benar.
 - Bahasanya singkat dan sesuai untuk kandidat.
 - Informasi yang cukup untuk menjawab pertanyaan.
 - Hanya ada jawaban yang benar.
 - Pengganggu adalah kesalahan tetapi masuk akal dan mendiskriminasikan pas level yang benar.
 - Pilihannya homogen, panjangnya sama, tidak ada sangkut pautnya, dan soal itu sesuai untuk test.
4. Menghasilkan banyak waktu dan persyaratan untuk memperoleh sejumlah soal yang memuaskan yang diharuskan untuk bagian. Khususnya untuk menguji keterampilan seperti berselancar. Masalah yang khusus terletak pada penentuan jawaban pengecoh yang sesuai bagi soal-soal yang menguji keterampilan reseptif yang lebih luas. Heuston (1975) mencatat bahwa lebih berguna membuat pertanyaan-pertanyaan open-ended daripada soal-soal pilihan ganda; sebaliknya para siswa harus mengingat 4 atau 5 pilihan bagi tiap soal sementara mereka mencoba untuk mengolah teks.
 5. Keberatan berikutnya terhadap penggunaan pilihan ganda yaitu bahaya dari format yang mempunyai efek yang tidak seharusnya terhadap penilaian sifat. Ini sudah menjadi bukti bagi hubungan yang rendah dengan penilaian membaca alternatif dan dengan data validitas eksternan concurren lainnya pada kemampuan membaca para kandidat (lihat Weir, 1983a).
 6. Ada keraguan yang benar tentang validitas mereka seperti penilaian kemampuan bahasa; menjawab soal pilihan ganda merupakan sebuah tugas yang tidak nyata, seperti dalam kehidupan nyata, seseorang jarang dihadapkan dengan empat pilihan untuk membuat pilihan pemahaman yang baik.

Normalnya, ketika diperlukan suatu pemahaman terhadap apa yang telah dibaca atau didengar bisa disampaikan lewat ucapan atau tulisan. Dalam test pilihan ganda, pengganggu memberikan pilihan untuk berpikir, tapi mungkin malah sebaliknya memberikan kesempatan untuk tidak berpikir. Jika ada pandangan yang berbeda mungkin dibantah bahwa kadang-kadang ada lebih dari satu jawaban yang benar untuk beberapa pertanyaan, khusus pada tingkat akhir. Apa yang penyusun test duga sebagai jawaban yang benarmungkin tidak bagi apa yang lainnya, yaitu harus jelas dalam sebuah teks.

5.1.2 Pertanyaan-pertanyaan Jawaban Singkat

Ini merupakan pertanyaan-pertanyaan di mana para kandidat harus mencatat jawaban yang spesifik pada tempat yang disediakan dalam lembar jawaban. Teknik ini benar-benar berguna untuk ujian pemahaman membaca dan menyimak. Pendapat-pendapat di bawah ini berkenaan dengan membaca juga dapat dipakai pada ujian menyimak.

Manfaat

1. Jawaban tidak disediakan seperti pada pilihan ganda; oleh karena itu jika seorang siswa menjawab benar, itu tidak akan terjadi kecuali ia memahami teks.
2. Dengan rumusan pertanyaan yang teliti, jawaban seorang kandidat bisa singkat dan oleh karena itu sebagian besar pertanyaan mungkin disusun dalam format ini.
3. Jika beberapa jawaban yang dapat diterima dibatasi, memungkinkan untuk memberikan instruksi yang benar kepada para penguji yang menilainya.
4. Aktivitas-aktivitas seperti kesimpulan, pengenalan serangkaian, perbandingan dan penentuan ide pokok suatu teks, ini bisa dilakukan dengan efektif melalui pertanyaan-pertanyaan jawaban singkat dimana jawaban harus dicari dari teks yang tersedia.
5. Sebuah kasus yang kuat bisa dibuat konteks yang sesuai, contohnya pada test EAP, untuk penggunaan teks yang panjang dengan format jawaban singkat

dibanding yang lebih representatif pada bacaan yang diperlukan dalam situasi target. Mereka juga bisa memberikan data yang lebih dapat dipercaya tentang kemampuan membaca seorang calon (lihat Engineer, 1977, untuk bukti tentang kemampuan yang meningkat hasil dari penggunaan teks yang panjang dan Appendix I contoh dari pendekatan ini pada test TEEP).

Kerugian

1. *Kerugian* utama pada teknik ini adalah bahwa ia menyebabkan kandidat menulis dan ada beberapa soal yang mengganggu penilaian konsep yang diharapkan.
2. Perhatian diperlukan dalam penyusunan soal-soal untuk membatasi jawaban-jawaban yang mungkin dapat diterima dan banyak tulisan yang diperlukan. Dalam kasus-kasus tersebut dimana terdapat banyak pembahasan tentang hal yang dapat diterima sebuah jawaban, contohnya, pada pertanyaan-pertanyaan yang memerlukan keterampilan-keterampilan menarik kesimpulan, ada sebuah kemungkinan di mana ketidaktepatan jawaban mungkin menuntun pada ketidak-dapat dipercaya penilai. Bagaimanapun juga, sikap tidak berlebihan dan standarisasi yang teliti dari penguji akan membantu mengurangi ini.

5.1.3 *Cloze*

Dalam prosedur *cloze*, kata-kata dihilangkan dari sebuah teks setelah pengenalan kalimat. Dasar penghilangan disusun dengan mesin, biasanya antara kata kel-5 dan ke-7. Para peserta harus mengisi tiap yang dikosongkan/celah dengan mengisi kata yang mereka pikir telah dihilangkan. Penelitian Alderson (1978a) membuktikan bahwa teks yang lebih sulit merupakan langkah yang lebih baik bagi keterampilan yang lebih rendah daripada teks yang mudah. Dia menemukan prosedur yang dapat diterima secara semantik akan menjadi yang lebih unggul dari yang lain.

Dalam perbandingan antara *cloze* dan pilihan ganda, Enginer (1977) memutuskan bahwa teknik-teknik ini mengukur aspek-aspek yang berbeda dari

aktivitas membaca artinya bahwa *cloze* mengukur ‘proses’ membaca, yakni kemampuan pembaca untuk memahami teks ketika ia membacanya; sebaliknya pilihan ganda mengukur hasil membaca-artinya kemampuan pembaca dalam menerjemahkan informasi yang abstrak terhadap nilai maknanya.

Ada satu kesepakatan yang baik dari bukti yang supportif untuk menggunakan bentuk *cloze*. (Klein-Braley, 1981, h.229) berpendapat bahwa: ‘sekarang ini, hasil penelitian menggunakan test *cloze* benar-benar memberi harapan. Mereka menunjukkan validitas yang tinggi, reliabilitas, objektivitas, diskriminasi, dan lain-lain yang tinggi’. Dia mengutip J.D. Brown (1979), ‘sebagaimana ditunjukkan dalam studi ini dan studi lainnya, ia bisa menjadi test yang valid dan relabel pada kemampuan bahasa yang kedua.’

Alderson (1978a, h.2) menggambarkan bagaimana: ‘dekade terakhir, memperlihatkan peningkatan penggunaan *cloze* pada bukan pengguna asli bahasa Inggris untuk mengukur tidak hanya kemampuan pemahaman membaca tetapi juga kemampuan linguistik umum mereka dalam bahasa Inggris yang merupakan bahasa Asing.’ Dia menambahkan (h.39):

“konsensus umum terhadap studi-studi dengan prosedur *cloze* pada 20 tahun terakhir telah menjadi ukuran yang dapat dipercaya dan valid bagi pemahaman membaca, bagi para pengguna asli bahasa Inggris.... Sebagaimana tindakan pemahaman teks, *cloze* telah ditunjukkan untuk membuat hubungan baik dengan jenis-jenis test lain dengan teks yang sama dan juga dengan ujian pemahaman membaca yang dibakukan.”

Dia menjelaskan bahwa meskipun bukti ini tidak bagi *non-native speaker* ‘tampaknya prosedur *cloze* merupakan langkah yang menarik bagi kemampuan bahasa bagi *non-native speaker*.’

Istilah *cloze* diperkenalkan pertama kali oleh W.L. Taylor (1953) yang mengambilnya dari konsep umum dari *closure* yang mengarah pada kecenderungan individu untuk menyesuaikan pola di mana mereka telah memegang arti seluruhnya. Taylor (h.416) menggambarkan sebagai berikut: ‘suatu unit *cloze* mungkin didefinisikan sebagai: beberapa peristiwa dari usaha untuk meniru keakuratan pada bagian yang dihilangkan dari sebuah “pesan” (beberapa produk bahasa), dengan menentukan bagian apa yang hilang itu’. Pembaca

memahami kalimat yang terpotong seluruhnya dan menyelesaikannya. Alderson (1978, h.8) menjelaskan bahwa ‘prosedur *cloze* menjadi ukuran kesamaan antara pola-pola yang diantisipasi oleh ahli sandi yang digunakan oleh penulis dalam sandi’.

Taylor pertama kali menggunakan prosedur ini untuk mengukur sejauh mana suatu teks dapat dibaca kemudian itu diharapkan bisa mengukur ujian pemahaman membaca dan kukan mengukur seluruh kemampuan bahasa. Bagi Bormuth (1962, h.134) ‘test *cloze* merupakan pengukuran-pengukuran yang sama-sama valid atas kemampuan pemahaman membaca.’ Heaton (1975, h.22) berpikir bahwa: ‘test *cloze* mengukur kemampuan pembaca untuk membaca sandi pada pesan-pesan yang terpecahkan dengan membuat penggantian-penggantian yang paling dapat diterima dari semua petunjuk kontekstual yang ada.’

Engineer (1977) menemukan bahwa test *cloze* yang diberikan di bawah kondisi waktu yang diatur. Memberikan indeks yang benar dan dapat dipercaya pada kemampuan para siswa. Jika dua kondisi terpenuhi: pertama, materi tekstual yang digunakan memiliki tingkat kesulitan yang sesuai dengan populasi, dan kedua, berisi sejumlah soal yang dihilangkan.

Manfaat

1. Test *cloze* mudah dalam menyusunnya dan menilainya. Jika prosedur yang menilai kata yang tepat digunakan. Mereka dituntut untuk menjadi indikator yang valid bagi seluruh kemampuan bahasa (lihat. Bormuth, 1962; Brown, 1979; Engineer, 1977; dan Oller, 1979).
2. Dengan penghilangan kata kelima pada sejumlah besar soal bisa disusun pada teks yang relatif pendek dan dapat menunjukkan tingkat konsistensi internal yang tinggi. Konsistensi ini mungkin berubah-ubah, mekipun, terikat pada teks yang dipilih, memulai penghilangan dan dasar penghilangan yang dipakai.
3. Dalam literatur test-test *cloze* sering dianggap sebagai ukuran-ukuran yang benar dan sama pemahaman membaca.

Kerugian

1. Meskipun argumen-argumen yang dikemukakan untuk kepentingan prosedur *cloze*, beberapa keraguan telah terungkap. Sebagian besar mengenai validitasnya sebagai perlengkapan ujian.
2. Alderson (1978, h.392) mengemukakan bahwa:
Prosedur *cloze* bukan prosedur kesatuan, sejak ada satu kekurangan yang dinilai dari sifat yang bisa dibandingkan diantara test mungkin digunakan. Fakta membuktikan dengan jelas bahwa test-test *cloze* yang berbeda, dihasilkan oleh beberapa variasi khususnya dari faktor yang tidak tetap, memberikan ukuran-ukuran yang berbeda yang tidak bisa diperkirakan, khususnya pada kemampuan berbahasa Inggris sebagai bahasa Asing.
Jika seseorang mengubah teks, mengubah dasar penghilangan, mulai dari tempat yang berbeda atau mengubah prosedur penilaian, seseorang yang mendapatkan test yang berbeda mengenai koefisien reliabilitas dan validitas dan seluruh kesulitan test.
3. Bukti ini bertentangan dengan cara membedakan metode-metode penilaian yang akan dipakai dalam menilai prosedur *cloze*. Klein-Braley (1985) memberi kesan bahwa test *cloze* merupakan pengukuran yang sangat kurang baik dengan ukuran-ukuran kemampuan umum lain yang ditentukan ketika digunakan pada satu bahasa seperti pada berbagai bahasa. Dan itu tampak bahwa *cloze* tidak sesuai dengan kelompok yang terbatas (Klein-Braley, 1985); hubungan yang lemah telah ditemukan antara *cloze* dan pertimbangan guru (Klein-Braley, 1981; 1985); *cloze* tidak tampak hubungan baik dengan test-test produktif berbicara dan menulis dan skor-skor pada *cloze* tidak bisa dihubungkan dengan *native speaker* dengan mudah sejak performa *native speaker* berubah dari suatu test *cloze* ke test yang lain (Alderson, 1978a).
4. Prosedur *cloze* terlihat menghasilkan test-test yang lebih berhasil bagi sintaksis dan leksik dan level kalimat daripada bagi pemahaman membaca pada kemampuan umum atau inferensial atau deduktif, apa yang mungkin

disebut dengan kemampuan yang lebih layak (lihat Darnell, 1968). Ini akan terlihat sesuai dengan pendapat Alderson (1978, h.99) bahwa:

“pada dasarnya *cloze* merupakan loncatan kalimat... jelasnya, fakta bahwa prosedur *cloze* menghilangkan beberapa kata daripada frase atau klausa harus membatasi kemampuannya akan pemahaman test lebih banyak daripada terhadap lingkungan dekatnya, sejak kata-kata individu tidak mempengaruhi kepaduan tekstual dan hubungan percakapan (dengan pengecualian yang nyata dan alat-alat yang bersatu padu seperti anaphora, leksikal, pengulangan, dan penghubung yang logis).”

5. Mungkin syarat yang paling penting yaitu pertanyaan tentang performa apa yang ada dalam test *cloze* yang benar memberi tahu kita tentang kemampuan bahasa peserta. Sulit untuk menafsirkan skor test *cloze* ke dalam deskripsi tentang apa yang bisa dilakukan oleh peserta atau apa yang tidak bisa dilakukan dalam kehidupan nyata.

5.1.4 Mengisi Celah Penghilangan Selektif

Mengingat penemuan-penemuan negatif baru-baru ini mengenai penghilangan *cloze*, pemilihan soal-soal untuk penghilangan berdasarkan apa yang diketahui mengenai bahasa, mengenai kesulitan dalam memahami teks, yang diketahui cara bahasam mengenai kesulitan dalam memahami teks, mengenai cara bahasa bekerja dalam teks-teks tertentu. Pertimbangan linguistik digunakan untuk melakukan penghilangan dan maka lebih mudah untuk meneruskan apa yang diharapkan tiap test untuk mengukur (lihat Alderson, 1987a, h.397; Klein-Braley, 1981, h.244; dan Weir, 1983a). teknik ini disebut sebagai mengisi celah penghilangan selektif bukan *cloze*.

Manfaat

1. Penghilangan selektif memungkinkan pembina test untuk memutuskan di mana penghilangan dibuat dan memutuskan pada soal-soal yang telah diseleksi berdasarkan teori dan menjadi penting bagi para peserta target tertentu.

2. Juga mudah bagi penulis test untuk membuat beberapa perubahan menunjukkan untuk menajadi analisis soal yang penting dan mempertahankan beberapa soal yang dikehendaki. Ia mungkin melibatkan penghapusan soal yang be;um memuaskan perihal membedakan nilai kecakapan.

Kerugian

1. Ini penting untuk menekankan bahwa tehnik ini membatasi satu untuk penarikan contoh pada keterampilan-keterampilan yang memungkinkan (yakni kemampuan-kemampuan yang secara kolektif menggambarkan keterampilan membaca) daripada melaksanakan format jawaban singkat atau pilihan ganda (lihat Weir). Sedangkan pertanyaan-pertanyaan jawaban singkat dan pilihan ganda membolehkan penarikan contoh keterampilan-keterampilan membaca lebih banyak, mengisi celah lebih terbatas dimana hnaya satu kata yang dihilangkan.
2. Jika tujuan suatu test untuk mencoba keterampilan yang memungkinkan termasuk keterampilan yang lebihluas seperti berselancar, maka format tamahan intik 'mencari celah' diperlukan sekali.

5.1.5 C-Test

Baru-baru ini sebuah alternatif bagi *cloze* dan mencari celah penghilangan selektif telah muncul untuk test pemahaman pada unsur-unsur linguistik yang lebih spesifik dalam sebuah teks. Penyesuaian teknik *cloze* yang disebut C-Test yang telah dikembangkan di jerman oleh Klei-Braley (1981, 1985; Klein-Braley dan Raatz, 1984) yang berdasarkan pada dasar pemikiran teoritis yang sama seperti *cloze*, viz., yang menguji kemampuan untuk mengatasi pleonasme (kelebihan) yang direduksi dan membuat ramalan dari konteks.

Dalam C-Test setiap kata kedua di dalam sebuah teks dihilangkan sebagian. Dalam usaha untuk memastikan solusi para siswa diberi setengah dari kata pertama yang dihilangkan. Peserta ujian menyelesaikan kata di atas kertas ujian dan prosedur penilaian satu kata yang tepat dilakukan.

Manfaat

1. Dengan menggunakan C-Test variasi teks dianjurkan, dan diberikan soal dalam jumlah besar yang bisa disebabkan oleh teks yang sederhana, selanjutnya ini mempertinggi sifat dasar yang representatif dari bahasa yang sedang dicoba. Normalnya minimal 100 penghilangan yang dibuat dan lebih representatis pada sebagian daripada yang mungkin di bawah teknik *cloze*.
2. Tugas bisa dinilai secara objektif, karena jarang ada lebih dari satu jawaban yang memungkinkan dari satu celah.
3. Sedangkan dalam *cloze* performa *native speaker* pada test merupakan faktor yang tidak tetap, menurut Klein-Braley (1985) lebih lazim bagi *native speaker* untuk bisa mendapatkan skor 100% pada C-Test. Mungkin karena beberapa bantuan dalam mengurangi skor, contohnya apa keuntungan mendapat nilai.
4. C-Test hemat dan hasil yang diperoleh sekarang mendorong reliabilitas dan validitas internal dan eksternal. Akan terlihat menggambarkan alternatif yang aktif pada prosedur *cloze* dan *selective deletion gap filling*.

Kerugian

1. Memperlihatkan secara relatif teknik dalam bentuk ini terdapat bukti yang sedikit empiris dari nilainya. Sebagian besar perhatian telah diberikan mengenai yang dapat diterima umum sebagai ukuran kemampuan bahasa. Menari mengetahui bahwa Davies (1965) memiliki versi dari teknik ini dalam rentetannya di mana tulisan pertama satu kata diberikan.
2. Teknik ini cacat dari fakta bahwa ia menjengkelkan bagi para siswa yang harus mengolah teks yang rusak dengan keras dan validitas luar prosedur rendah.

4.1.6 *Cloze Elide*

Sebuah teknik yang menarik yaitu di mana kata-kata yang tidak semestinya disisipkan ke dalam bacaan dan para peserta harus menunjukkan di mana letak sisipan tersebut. Kenyataannya, tidak ada yang baru mengenai teknik ini, Davies menggunakannya baru-baru ini (Davies, 1985). Dalam bentuknya yang lebih baru, ia dikenal sebagai teknik *ata-kata yang kacau*.

Manfaat

1. Pada perbandingan dengan format pilihan ganda atau pertanyaan jawaban singkat, peserta tidak bermasalah dalam memahami pertanyaan. Kira-kira ia mempunyai persamaan, ia disebut sebagai test *cloze*.

Kerugian

1. Penilaiannya bermasalah karena peserta mungkin mencoret soal-soal yang benar, tetapi berlebih-lebihan.

4.1.7 Transfer Informasi

Dalam ujian baik pemahaman membaca maupun menyimak kita telah dihubungkan dengan masalah pengukuran yang menjadi '*muddied*' dengan harus menggunakan tulisan terhadap jawaban catatan. Usaha untuk menghindari kontaminasi pada skor-skor beberapa papan ujian di Inggris telah memasukkan tugas di mana informasi yang dikirimkan secara verbal ditransfer menjadi bentuk non-verbal, contohnya dengan mengisi diagram, menyelesaikan grafik, atau mengurutkan peristiwa (lihat Appendix V untuk contoh yang menarik dalam test JMB).

Manfaat

1. Teknik transfer informasi cocok untuk ujian pemahaman proses, klasifikasi atau urutan narasi dan berguna untuk ujian variasi jenis teks yang lain. Ia menghindari kemungkinan kontaminasi dari para siswa yang harus mengisi penuh jawaban.

2. Itu merupakan tugas yang realistis untuk situasi-situasi yang berbeda dan minatnya dan keaslian memberikannya validitas luar dalam konteks ini.

Kerugian

1. Banyak sekali perhatian yang harus diberikan tugas non-verbal itu peristiwa harus menyelesaikan proses. Pada beberapa tugas para siswa mungkin harus memahami teks.
2. Ada bahaya bias budaya dan pendidikan. Pada subjek tertentu para siswa mungkin juga diragukan, contohnya, beberapa siswa kelas sosial mungkin tidak mahir dalam mengerjakannya dalam medium non-verbal seperti teman-temannya di kelas IPA.

Kesimpulan

Untuk ujian kemampuan membaca kami akan menganjurkan penggunaan soal-soal jawaban singkat dan mengisi celah hilang selektif secara bersamaan. C-Test merupakan alternatif yang menarik berikutnya dan hal yang dapat dipercaya para siswa yang validitasnya berguna bagi investigasi selanjutnya. Jika kita harus mengembangkan sifat dasar yang komunikatif dari test-test kita, mungkin penting untuk memusatkan pada tugas-tugas performa di dalam test-test membaca, dan penggunaan teknik-teknik transfer informasi dan format-format respon terbatas yang dianjurkan.

5.2 Ujian Menyimak Pemahaman

5.2.1 Ujian Kemampuan Menyimak Secara Ekstensif

Dasar pemikiran di belakang konstruksi beberapa test pemahaman menyimak sekarang ini dijelaskan oleh Valette (1967, h.49): ‘objek pokok test menyimak adalah evaluasi pemahaman siswa. Tingkat pemahamannya akan tergantung pada kemampuannya membedakan fonem, mengenal tekanan dan intonasi, dan memahami apa yang ia dengar.’

Terpikir bahwajika seorang pelajar diuji pada perbedaan fonem, tekanan dan intonasi, jumlah sub-test yang ‘berbeda’ akan sama dengan kemampuannya dalam menyimak. Contoh test jenis ini adalah rangkaian test ELBA yang dikonsepsi oleh Ingram (1964) yang menekankan pada soal-soal menyimak yang ‘berbeda’ seperti pengenalan suara, intonasi, dan tekanan, penggunaan soal-soal singkat daripada penggunaan percakapan atau dialog yang terus-menerus. Sebagaimana ryan (1979) jelaskan, bahkan bagian yang disebut sebagai pemahaman menyimak terlihat lebih banyak merupakan test mekanisme respons yang sesuai dengan test pemahaman percakapan yang terus menerus pada konteks asli.

Kecenderungan yang nyata di tahun-tahun ini adalah usaha membedakan antara test-test diskriminasi yang berhubungan dengan pendekatan dan test-test yang berhubungan dengan konteks pemahaman menyimak. Templeton (1973) menjelaskan bagaimana penelitian mulai memusatkan pada test-test integratif. Pemahaman menyimak ini yang merujuk pada test-test poin berbeda pada diskriminasi fonem, intonasi dan kata, dan tekanan kalimat.

Sejak 1969 JMB bukan lagi merupakan test-test keterampilan yang berhubungan dengan pendengaran individu yang terisolasi, tetapi merupakan test pemahaman menyimak pada gabungan konteks ceramah atau dialog (lihat McEldowney, 1976, dan Appendix IV). Perubahan paradigma ini juga bisa diobservasi dengan versi EPTB tahun 1977 (lihat Davies, 1978) yang menggantikan tugas-tugas analisis, diskriminasi fonem, tekanan, dan intonasi, seluruh muatan sub-test pemahaman menyimak, contohnya, test gabungan dari pemahaman menyimak yang berdasarkan pada ceramah dengan mengambil catatan yang disimulasikan.

Davies (1978, h.16-8) menjelaskan bagaimana perubahan-perubahan yang sama yang terjadi antar tugas-tugas menyimak yang dijelaskan dalam buku Valette (1976) dan Valette (1977) sebagai perubahan dari linguistik ke sosiolinguistik, dari strukturalisme menuju fungsionalisme, dari taksonomi dan perincian ke dalam keterampilan, kedalam bagian-bagian yang berbeda, integrasi dan penambahan menjadi menyeluruh. ‘Dalam Valette edisi kedua (1977) Davies

berpendapat (h.147): ‘perubahan dari memusatkan pada suara, hasil percakapan, fonologi, kedalam makna dan komunikasi.’

Argumen kuat yang menentang diskriminasi yang berhubungan dengan pendengaran sebagai test kemampuan pada pemahaman menyimak adalah bahwa kemampuan untuk membedakan antara fonem-fonem meskipun tidak menyatakan kemampuan untuk memahami pesan verbal. Bagi Valette (1977, h.102): ‘perhatian khusus penguji yaitu untuk mengetahui apakah para siswa menerima pesan yang dimaksud dan bukan pada pembuatan diskriminasi suara tertentu atau identifikasi ciri-ciri struktural tertentu.

J.W. Morrison (1974) setelah menilai pemahaman menyimak, ia menyimpulkan bahwa dalam test ESP performa komunikatif harus diertimbangkan pada level yang melebihi test fonologi dan gramatikal, kemudian membuat laporan mengenai konteks komunikatif percakapan. Chaplen (1970, h.19) baru-baru ini menyimpulkan bahwa: ‘apapun kontribusi dari unsur-unsur komunikasi-intonasi, tekanan, diskriminasi fonemik-terhadap test komunikasi, kepenringan mereka terlihat paling rendah pada beberapa tingkat kecakapan di tingkat dasar.’

Holes (1972) mengembangkan instrumen test yang memfokuskan pada kemampuan mengenai ceramah-ceramahnya.

a. Soal-soal Pilihan Ganda

Pada pertimbangan kita mengenai penggunaan teknik ini dalam penilaian yang telah dibahas pada 4.1.1, jelas bahwa *Kerugian* menggunakan test ini diikuti oleh beberapa *Manfaat* yang mungkin ia miliki.

Karena masalah-masalah yang berhubungan dengan sifat dasar utama dari proses menyimak menyebabkan adanya kesulitan-kesulitan tambahan dalam menggunakan tehnik ini seperti ukuran kemampuan menyimak, contohnya beban tambahan yang diberikan pada pengolahan dengan harus mengingat empat pilihan (Heaton, 1975). Formatnya dibuat dan makin terasa sebagai metode yang invalid untuk menilai pemahaman oleh para guru, para perencanaan materi dan para penguji bahasa. Sertifikat umum yang baru dari ujian pendidikan menengah

(GCSE) di Inggris tidak akan menggunakan format pilihan ganda secara besar dikarenakan komentar pedas dari organisasi guru atas validitasnya sebagai pengajaran dan ujian.

Ujian RSCA CUEFL, meskipun upaya-upaya untuk memaksimalkan keaslian pada teks pendukung diseleksi, benar-benar tergantung pada format ini dan perbedaannya (contohnya, soal-soal benar-salah) dan telah dikritikkarena kemundurannya dari pengolahan percakapan realistik yang terlibat (lihat Appendix III). Penggunaannya oleh RSA atas format-format yang lebih objektif ini menyoroti perlunya mencoba menentukan realisme stimulus teks dan tugas yang diharapkan bisa diberikan pada siswa, dan kadang-kadang tidak sesuai dengan reliabilitasnya dan validitas.

2. Soal-soal Jawaban Singkat (SAQs)

Manfaat

1. Soal-soal jawaban singkat bisa menjadi aktivitas yang realistik bagi ujian pemahaman menyimak, contohnya, jika seseorang berharap menyimulasikan aktivitas-aktivitas kehidupan nyata seseorang menyampaikan pesannya. Dengan perhatian yang cukup, jawaban bisa dibatasi dan kemudian bahaya proses 'menulis' yang mengganggu pengukuan menyimak sebagian besar dihindari (lihat Appendix I).
2. Berbeda dengan format-format pilihan ganda dan benar salah yang digunakan beberapa ujian, seseorang bisa menjadi lebih khusus yaitu bahwa jawaban-jawaban yang benar tidak datang secara kebetulan.

Kerugian

1. Jika peserta harus menulis jawaban pada waktu yang sama seperti menyimak percakapan yang terus-menerus maka terdapat masalah-masalah yang nyata. Muatan yang tidak penting mungkin dimasukkan kedalam ingatan dan informasi yang sangat penting dari percakapan yang terus menerus mungkin hilang ketika jawaban soal sebelumnya sedang ditulis.

c. Teknik Transfer Informasi

Teknik ini sudah dibahas di atas yang berhubungan dengan membaca dan berguna untuk pertimbangan alasan-alasan yang sama pada test menyimak (lihat Appendix V).

Manfaat

1. Manfaat utama menggunakan teknik ini dalam ujian menyimak adalah bahwa siswa tidak harus mengolah soal-soal tertulis ketika mencoba pemakaian yang diucapkan yang masuk akal. Ini efisien khususnya untuk ujian sebuah pemahaman rangkaian proses yang berhubungan dengan teks dan klasifikasi.

Kerugian

1. Sangat sulit untuk menemukan teks-teks yang diucapkan yang menuntun dirinya sendiri pada format non-verbal. Sedangkan dalam 'membaca' sejumlah editing teks dapat dikerjakan dengan mudah dan dalam variasi teks yang lebih besar lebih tersedia dengan mudah, ini bukan merupakan kasus bagi teks menyimak dari sumber-sumber autentik.

5.2.1 Batasan-batasan untuk Ujian Menyimak Ekstensif

Perlu diketahui bahwa jika seseorang berharap membuat tugas-tugas test lebih mirip dengan yang ada dalam kehidupan nyata, sifat dasar percakapan yang diperpanjang dan masalah-masalah pengolahan yang lebih besar berhubungan dengan pemahaman soal-soal yang menghalangi ucapan bahasa Inggris yang memusatkan pada keterampilan linguistik yang lebih khusus seperti menentukan makna kata-kata dari konteks yang mengenalkan nilai makna dari ciri-ciri khusus tekanan atau intonasi. Benar-benar sulit bagi siswa untuk kembali dan memfokuskan pada ciri-ciri yang sangat spesifik dari percakapan ketika menyimak dan mencoba memahami percakapan non-interaktif, monolog yang terus menerus. Oleh karena itu, untuk mempertahankan sifat dasar integratif test, kita harus memusatkan perhatian pada soal-soal keterampilan mengolah yang

lebih umum seperti menyimpulkan, menyimak yang pokok-poko dan menentukan ide utama.

Masalah yang serius dalam ujian menyimak ekstensif dengan menggunakan tape recorder yaitu bahwa unsur visual, banyaknya referensi eksoporic biasa dan informasi paralinguistik yang lebih sulit bagi para kandidat. Biasanya pendengar tidak harus mengolah suara-suara yang tidak ada dari tape recorder di kehidupan nyata (selain dari pengecualian yang nyata seperti mendengarkan radio).

Sampai ada hal yang dapat masuk ke dalam peralatan video, kepalsuan tugas menyimak akan tetap menjadi masalah. Bahkan video mungkin memiliki kesulitan prakteknya sendiri meskipun beberapa *screen* mengharuskan semua penontonnya diperlukan sama dalam ketidakcocokan dengan sistem-sistem yang berbeda.

Ada suatu bahaya yang besar dalam test-test menyimak di mana para peserta mungkin diharapkan untuk memecahkan kesulitan-kesulitan lain yang muncul dari konteks terbatas yang ada dan langkah-langkah selanjutnya harus diambil untuk mengimbangi ini atau harus meremehkan kemampuan mengolah bahasa lisan dengan serius.

5.2.2 Ujian Menyimak secara Intensif

Kesulitan-kesulitan dalam memusatkan poin-poin menyimak yang khusus telah dibahas di atas, dimana para peserta ditunjukkan pada percakapan yang terus menerus perlu mempertinggi reliabilitas rangkaian test kita sebaiknya memasukan beberapa soal yang spesifik. Dikte atau mengingat tes yang didengar dapat memberikan perbedaan ini sebaik menjadi valid dalam isi untuk kelompok-kelompok peserta tertentu, khususnya yang menyangkut studi akademis melalui medium bahasa Inggris.

Dikte

Hal ini sangatlah penting mengingat para peserta akan dinilai setelah mungkin dimana mereka akan diharuskan untuk menggunakan bahasa. Bagi dikte,

ini melibatkan, mereka menyimak materi yang didiktekan yang memasukkan pesan lisan yang khas mungkin mereka alami dalam situasi target.

Manfaat

1. Memperhatikan reliabilitas sebaik validitas, mungkin sebaliknya untuk memperbaiki seluruh rangkaian ‘menyimak’ dengan memasukan format yang memiliki laporan yang jarang terbukti pada respek ini. Dikte dapat memberi reliabilitas ini melalui beberapa soal yang bisa dihasilkan sebaik dijadikan valid bagi situasi-situasi tertentu di mana dikte mungkin menonjol seperti aktifitas kelompok target.
2. Ada beberapa bukti yang menunjukkan bahwa dikte berhubungan erat dengan variasi-variasi test lain, khususnya dengan test-test integratif lain seperti *cloze* dan sering digunakan sebagai ukuran kecakapan umum yang berguna. Ada beberapa bukti bahwa penggunaan skema penilaian semantik (lihat Weir, 1983a) seperti terhadap sistem kata yang tepat membantu mempertinggi hubungan dengan konsep test-test menyimak yang valid lainnya.
3. Kecaman atas dikte masa lalu berasal dari sudut pandang yang banyak sekali dipengaruhi oleh linguistik struktural yang mendukung uji unsur-unsur keterampilan bahasa yang lebih berbeda dan berharap mungkin dari kemungkinan pengukuran yang keruh. Heaton (1975) berkomentar: ‘seperti alat ujian, ia mengukur begitu banyak ciri-ciri bahasa yang berbeda untuk menjadi efektif dalam memberikan cara untuk menilai keterampilan khusus seseorang. Bagaimanapun juga, para pendukung dikte mempertimbangkan sifat dasarnya yang sangat ‘integratif’ agar ber*Manfaat* sejak ia menggambarkan dengan tepat bagaimana orang-orang mengolah bahasa dalam konteks kehidupan nyata.
4. Minat baru pada dikte menggambarkan perubahan paradigma dalam nilai-nilai ujian dan tujuan-tujuannya telah dibahas. Padahal pada tahun 1967 Valette telah meneliti bahwa para spesialis bahasa asing tidak setuju dengan keefektipan dikte sebagai suatu ujian beberapa siswa yang lebih maju, sepuluh tahun kemudian dia bisa menguraikan bahwa dikte merupakan

ukuran yang tepat bagi seluruh kecakapan dan metode yang luar biasa untuk mengelompokkan para siswa yang baru masuk berdasarkan tingkat kemampuannya.

5. Faktor yang penting kembalinya dikte pada popularitas sebagai alat ujian merupakan penelitian yang dilakukan oleh Oller, yang membentuk minat yang lebih besar terhadap ujian integratif. Oller (1979) menolak kecaman dikte dan membuktikan bahwa ia merupakan test pemahaman menyimak yang tepat karena ia menguji rangkaian keterampilan integratif yang luas.
6. Oller mengklaim bahwa proses analisis dinamis dengan sintesis itu berbelit-belit. Dikte menarik kemampuan pelajar untuk menggunakan semua sistem bahasa bersama dengan pengetahuan tentang dunia, konteks, dan lain-lain, untuk memperkirakan apa yang dikatakan (perpaduan pesan) dan setelah pesan diucapkan diperiksa dengan teliti lewat ingatan singkat untuk mengetahui apakah ia sesuai dengan apa yang telah dipekirakan.
7. Bagi Oller, dikte menguji tidak hanya kemampuan siswa untuk membedakan unit-unit fonologikal tetapi juga kemampuannya untuk membuat keputusan mengenai batas-batas kata; dengan cara ini seseorang yang diuji menemukan urutan kata-kata dan frase-frase dan dari ini ia membangun ulang suatu pesan. Identifikasi kata-kata dari konteks seperti dari suara-suara yang diterima terlihat oleh Oller sebagai *Manfaat* dikte yang positif maka kemampuan ini sangat penting dalam membuat bahasa menjadi berfungsi. Keberhasilan membangun ulang pesan oleh siswa dikatakan tergantung pada tingkat 'tatabahasa harapan' yang diinternalkannya meniru itu dari *native speaker*. *Native speaker* yang mahir slalu mendapat skor 100% dari dikte yang dikelola dengan baik sementara pelajar non-native membuat kesalahan-kesalahan pada penghilangan, sisipan, perintah kata, inversi, dan lain-lain, menunjukkan bahwa tatabahasa yang diinternalkan tidak akurat dan tidak sempurna; mereka tidak sepenuhnya mengerti apa yang mereka dengar dan apa yang mereka tulis ulang dalam studi yang berbeda-beda bersamaan dari yang asli.

8. Menurut Oller, penelitian menunjukkan bahwa hasil test dikte merupakan peramal kemampuan bahasa yang kuat daripada yang diukur oleh jenis-jenis test bahasa lain (lihat Oller, 1971; Valette, 1977).

Kerugian

1. Alderson (1978a) menyimpulkan bahwa bukti mengenai dikte tidak meyakinkan dan berguna hanya sebagai rangkaian test menyimak daripada sebagai solusi tunggal. Dia (1978a, h.365) menyatakan bahwa:
Alasannya adalah lebih banyak berhubungan dengan beberapa sub-test daripada dengan yang lainnya yang tidak terlihat menjadi fakta yang semestinya dikalim bahwa ia merupakan test integratif, tetapi karena secara esensial merupakan test keterampilan linguistik level awah. Oleh karena itu, dikte mempunyai hubungan paling baik dengan test-test *cloze*, teks-teks dan metode-metode penilaian yang mereka sendiri memperbolehkan pengukuran keterampilan-keterampilan ini.
2. Dikte akan menjadi hal sepele. Jika tidak memori siswa merupakan tantangan dan panjangnya kata-kata yang didiktekan akan tergantung pada kemampuan pendengar yang terbatas di man arekan-rekan *native speaker* bisa menanganinya.
3. Penilaian mungkin menjadi suatu masalah, jika seseorang berharap membawanya ke dalam keseriusan kesalahan atau jika seseorang berharap menggunakan skema penilaian yang dikenal secara lebih komunikatif ketika nilai diberikan jika peserta telah memahami isi pokok pesan dan ciri-ciri redundan (pleonastis) diabaikan.
4. Jika dikte tidak diberikan dengan menggunakan tape, test akan kurang reliabel, seperti akan ada perbedaan-perbedaan pada kecepatan pengiriman teks kepada para pendengar yang berbeda.
5. Latihan bisa menjadi tidak realistis jika teks yang digunakan telah dibuat sebelumnya untuk dibacakan daripada didengarkan.

Listening Recall

Berbeda dari dikte, bukti yang ada mengenai listening recall sangat sedikit (lihat Furneaux, 1982; dan Beretta 1983 untuk penjelasan yang lengkap mengenai prosedur ini). Siswa diberi salinan dari bagian kata-kata tertentu yang telah dihilangkan (penghilangan ini diperiksa terlebih dahulu untuk memastikan bahwa mereka tidak bisa diisi hanya dengan membacanya). Kata-kata yang dihilangkan biasanya kata-kata isi yang dirasa penting untuk memahami percakapan dan tempat yang dikosongkan itu terjadi pada interval yang sering.

Para siswa diberi sedikit waktu untuk membaca teks, memperbolehkan penggerakan tatabahasa yang diharapkan. Mereka harus mengisi titik-titik, setelah mendengarkan tape recorder bagian yang komplit dua kali. Pertama kali mereka dianjurkan untuk mendengarkan kemudian berusaha untuk mengisi titik-titik dalam waktu singkat. Mereka diperbolehkan untuk menulis jawaban-jawabannya. Mereka mendengarkan bagian yang kedua kali dan kemudian dalam waktu singkat untuk menulis beberapa kata yang hilang. Waktu menulis yang terbatas ini mendatangkan memori yang singkat. Format ini melibatkan beberapa faktor linguistik yang telah dibahas untuk dikte dan direfleksikan kedalam nama-nama lain yang telah diberikan seperti dikte spot, dan kombinasi *cloze* dan dikte.

Manfaat

1. Seperti dikte, ia bisa dilakukan dengan cepat dan dinilai secara objektif dan membiarkan penguji memutuskan pada soal-soal yang dianggap penting.
2. Hubungan erat dengan test menyimak yang lebih langsung lainnya (Beretta, 1983) dan dengan total test untuk rangkaian menyimak telah disampaikan.
3. Memiliki keuntungan untuk pelaksanaan ujian dalam skala besar yaitu mudah merencanakan, melaksanakan, dan menilainya.

Kerugian

1. Kesulitan menggunakan teknik ini terletak pada perumusan apa yang diujikan. Sebagaimana halnya satu kata yang dihilangkan ia mungkin tidak menguji apapun lebih banyak dari kemampuan untuk menyesuaikan suara-

suara dengan simbol-simbol dibentuk oleh kemampuan untuk membaca bagian yang ada celahnya.

2. Ia merupakan tugas yang tidak autentik dan melibatkan kemampuan membaca sebaik kemampuan menyimak. Konsep yang teliti diperlukan untuk memastikan bahwa para siswa tidak dapat mengisi titik-titik hanya dengan membaca bagian tanpa harus mendengarkan semuanya.
3. Memberikan korelasi yang tinggi yang telah diketahui antara *listening recall* dan dikte (lihat Furneaux, 1982; Beretta, 1983), dan kepraktisan padan kata dan reliabilitas, validitas potensial dikte yang lebih besar bagi kelompok-kelompok tertentu, contohnya, bagi para siswa yang belajar melalui medium bahas Inggris, mungkin menuntun pada referensi untuk dikte lebih dari *listening recall*.
4. Permasalahan yang mungkin terjadi dalam penilaian dengan pertimbangan diberikan kepada apapun daripada *spelling* yang tepat.

Kesimpulan

Test-test menyimak memungkinkan untuk memasukkan tugas performa yang asli. Suatu upaya akan dibuat untuk teknik-teknik transfer informasi yang tidak tergabung (lihat appendix V). Kita mungkin memasukkan soal-soal jawaban singkat dengan berguna dan pertimbangan bisa diberikan kepada dikte (lihat Appendix I).

Ujian Menulis

Dua pendekatan bagi penilaian kemampuan 'menulis' bisa digunakan. Pertama, menulis bisa dibagi kedalam tingkat-tingkat yang berbeda, contohnya, tata bahasa, kosa kata, ejaan, dan tanda baca, dan unsur-unsur ini bisa diuji secara terpisah dengan menggunakan test-test objektif. Kedua, tugas-tugas menulis yang diperluas lebih langsung dari jenis-jenis yang berbeda bisa dikonsepsi. Inisi semua akan menjadi validitas konsep, isi, luar, dan *wasback* yang lebih besar tetapi akan membutuhkan penilaian yang lebih subjektif.

5.3.1 Metode tidak Langsung untuk Menilai Kompetensi Linguistik

Pada bagian 4.1, kita menguji format-format *cloze*, mengisi celah hilang selektif, C-Test, dan menguraikan nilai dari teknik-teknik ini untuk ujian linguistik yang lebih spesifik, keterampilan membaca kalimat, *viz*, soal-soal yang memusatkan pada pemahaman kosa kata, struktur, atau alat kepaduan.

Baik kemampuan produktif maupun ketereampilan reseptif bisa dierinci kedalam tingkatan tatabahasa dan leksikal berdasarkan kerangka poin yang berbeda. MCEldowney (1974, h.8) menguraikan test silabus JMB pada bahasa Inggris (bahasa asing):

Untuk bisa melaksanakan keempat keterampilan ini (menyimak, membaca, berbicara, dan menulis) dengan fungsi yang berbeda, penting untuk bisa memanipulasi soal-soal dari ketiga tingkatan bahasa. Yaitu, untuk berkomunikasi harus mempunyai kosa kata yang cukup, untuk mengetahui penggunaan tatabahasa bahasa Inggris dan untuk dapat menangani suara-suara bahasa Inggris, tekanan, dan intonasi.

Test JMB dalam bahasa Inggris (bahasa asing), seperti memasukkan tugas-tugas ujian hasil tertulis juga mempunyai tugas-tugas yang menguji pengetahuan mengenai 'kosa kata produktif dasar' dan 'soal-soal grammatikal' (lihat Appedix v). permasalahan yang menghadapkan konsep test-test kosa kata bermacam-macam. Chaplen (1970an) yang membuat konsep sub test bagian kosa kata dari test-test JMB baru-baru ini mencatat dua bidang masalah utama:

1. Pemilihan soal-soal leksikal;
2. Metode-metode yang digunakan untuk menguji soal-soal leksikal.

Jika orang-orang yang diuji mempelajari variasi subjek yang berbeda, maka terdapat suatu masalah yang serius dalam pemilihan. Lebih banyak materi yang digeneralisasi, lebih sulit untuk menyeleksi tujuan ujian. Dalam bidang-bidang yang diujikan khusus, dimana terdapat sesuatu yang dapat diidentifikasi, register yang disetujui, seleksi lebih mudah tetapi masih tetap sulit.

Permasalahan lain yang terjadi pada bobot relatif yang akan diberikan kepada soal-soal yang dipilih dari materi-materi membaca yang akan datang seperti terhadap soal-soal yang mungkin digunakan dalam tugas-tugas menulis

yang luas. Apakah kita menguji kosa kata aktif atau pasif? Selanjutnya, bagaimana kita menentukan tingkat frekuensi dan pentingnya soal-soal leksikal yang diperluas untuk digunakan dalam test?

Permasalahan yang sama terjadi pada seleksi soal-soal grammatikal untuk dimasukkan kedalam test kompetensi linguistik yang tidak langsung. Penelitian kuantitatif terhadap pembuatan soal-soal struktural pada materi-materi tertulis reseptif dan produktif, populasi test harus mengatasinya pada situasi target yang akan datang. Yang lebih pragmatis, metode pengambilan keputusan yang subjektif terhadap soal-soal yang akan diberikan itu diperlukan. Tampak perlu untuk menguji muatan test yang ada dan isi buku pelajaran pada tingkat yang sama untuk menentukan apa yang para ahli di bidangnya anggap sebagai soal-soal yang cocok bagi populasi yang sama.

Juga terlihat ada suatu masalah dalam menyampaikan apa yang sedang diujikan dalam test tatabahasa pon yang berbeda ini. Akankah performa test pengetahuan grammatikal tidak langsung disampaikan menurut profil untuk membaca dan menulis? Teknik tidak langsung lainnya terbatas pada persoalan yang berkenaan dengan validitasnya bagi para peserta ujian dan para pengguna hasil test tersebut. Usaha yang menarik untuk memelihara keobjektifan dan ulasan pendekatan poin yang berbeda ketika mempertinggi validitas bisa ditemukan pada tugas editing pada buku kedua mengenai test TEEP (lihat Weir, 1983a; dan Appendix I).

Tugas Editing

Dalam tugas editing siswa diberi teks yang berisi sejumlah kesalahan tatabahasa, ejaan, tanda baca, dari jenis-jenis yang ditulis biasa oleh para guru remedial bagi para siswa dalam kelompok target dan dimintai untuk menulis ulang bagian yang membuat semua koraksi penting.

Manfaat

1. Seperti pengukuran kompetensi yang lebih objektif, tugas ini mungkin memiliki efek washback yang baik di mana para siswa diajar dan didorong untuk memperbaiki tugas tertulis mereka dengan lebih teliti.
2. Pasti lebih banya *face valid* daripada teknik tidak langsung lainnya seperti menyesuaikan bagian proses menulis.

Kerugian

1. Jika siswa menulis ulang bagian-bagian kalimat dengan menggunakan kata-katanya sendiri daripada hanya mengoreksi kesalahan-kesalahan, masalah-masalah penilaian menjadi pertimbangan. Juga terdapat beberapa keraguan seperti apakah kemampuan untuk mengoreksi kesalahan-kesalahan orang lain sama dengan mengoreksi kesalahan-kesalahannya sendiri.
2. Penilaian juga bisamenjadi problematis jika peserta mengubah sesuatu yang sudah dikoreksi, sebuah kesalahan panitia bukan kelalaian.

5.3.2 Ujian Menulis Langsung

Dengan pendekatan yang lebih integratif dan langsung atas ujian menulis, kita bisa menggabungkan soal-soal yang menguji kemampuan peserta untuk mengerjakan tugas-tugas fungsional yang diperlakukan dalam performa tugas dalam situasi target. Bagi para dokter disebuah rumah sakit ini mungkin menyangkut penulisan surat kepada GP lokal mengenai seorang pasien. Bagi seorang siswa dalam konteks EAP mungkin menyangkut pencarian teks akademi untuk mengutip informasi yang spesifik yang akan digunakan pada rangkuman tertulis (lihat Appendix).

Test-test Essay

Ini merupakan sebuah metode tradisional bagi para siswa untuk membuat contoh tulisan yang berhubungan. Stimulus biasanya ditulis dan bisa mengubah panjangnya dari beberapa tulisan yang terbatas menjadi beberapa kalimat. Topik-topiknya seringkali sangat umum dan tergantung pada peserta yang mengeluarkan

isi kepalanya. Para peserta bagaimanapun juga tidak dituntun seperti bagaimana mereka diharapkan untuk menjawab pertanyaan.

Manfaat

1. Secara tradisional, essay telah diakui sebagai teknik ujian yang mungkin menjelaskan keengganan yang besar untuk membuangnya meskipun permasalahan dalam penilaian telah ditemukan (lihat Coffman, 1971; Gips dan Ewen, 1974).
2. Topik-topiknya benar-benar mudah dan merupakan teknik ujian yang terkenal bagi para peserta maupun bagi para pengguna hasil test. Maka ia memiliki validitas luar yang dangkal bagi orang awam.
3. Merupakan sarana yang cocok untuk menguji keterampilan-keterampilan, seperti kemampuan untuk mengembangkan argumen dengan cara yang logis, yang tidak bisa diuji dengan cara lain.
4. Manfaat besar sama dengan test-test menulis yaitu bahwa penulisan contoh dibuat yang bisa memberikan poin referensi bagi perbandingan yang akan datang.

Kerugian

1. Bebas, *open-ended writing* merupakan suatu masalah. Kemampuan untuk menulis topik-topik *open-ended* yang biasa mungkin tergantung pada latar belakang dan pengetahuan kultural peserta, khayalan atau kreativitas. Ini semua mungkin bukan merupakan faktor-faktor yang kita harapkan untuk dinilai.
2. Peserta mungkin tidak berminat dengan topik yang diberikan dan jika pemilihan topik yang diberikan sangat sulit untuk membandingkan performa khususnya jika hasil dari jenis-jenis yang berbeda dimasukkan.
3. Tekanan waktu sering menjadi ketidakleluasaan yang tidak realistis bagi essay dan tidak dilakukan di luar kehidupan akademis. Bagi sebagian besar orang proses menulis itu lebih lama dan mungkin melibatkan beberapa draft sebelum versi yang sempurna dibuat.

4. Pencantuman komponen ‘menulis’ pada suatu ujian menghabiskan waktu dari total waktu test yang ada untuk menguji semua keterampilan.

Tugas-tugas Menulis yang Terkontrol

Ternyata ada kasusus yang sangat kuat untuk memasukkan test menulis kedalam bidang validitas isi dari tugas-tugas ‘contoh tugas’. Ia menguji keterampilan penting yang tidak ada bentuk penilaian yang lain bisa dibuat contoh dengan tepat. Mengabaikan tugas ‘menulis’ pada situasi dimana tugas-tugas ‘menulis’ merupakan suatu ciri penting dari kebutuhan kehidupan nyata siswa mungkin benar-benar mengurangi validitas program uji.

Wall (1982) mengadakan investigasi yang memperjelas jenis-jenis tugas ‘menulis’ siswa teknik mesin diharuskan untuk menunjukkan bagian dari tugas sekolah mereka dan membandingkannya dengan tipe-tipe essay yang mereka set dalam rangkaian Michigan yang digunakan untuk menilai kecakapn bahasa siswa untuk masuk ke universitas. Dia (h.166) membuat kesimpulan mengenai perbedaan tersebut sebagai berikut:

Perbedaan utama terlihat bahwa dalam tugas-tugas ahli mesin ada banyak input berdasarkan teori dan tugas itu sendiri diuraikan dengan tegas, sedangkan dalam komposisi penulis hanya mempunyai harapan untuk merespons dan tidak harus hanya membuat isi tulisan tetapi juga membuat konteks, audiensi, dan tujuan dengan baik. Kriteria untuk menilai juga akan terlihat berbeda.

Kesimpulan harus investigasi itu menunggu penelitian yang menghasilkan: ‘korelasi studi antara total rangkaian Michigan dan bagian skor dengan GPA pertama siswa, yang tidak ada hubungan yang signifikan antara test dan kriteria guna kesuksesan akademis bisa ditemukan antara performa menulis dalam test dan indikator-indikator performa berikutnya dalam bida studi.

Menulis yang tidak terkontrol terlihat akan seperti test kemampuan ‘menulis’ yang tidak benar yang diperlukan oleh kebanyakan siswa. Lebih mudah untuk memperhitungkan kemungkinan dari test-test menulis ketika memperhatikan penentuan tiap tugas: media, peserta, tujuan, dan situasi, dengan aktivitas-aktivitas performa tingkat target (lihat Wall, 1982). Ketika tugas

ditentukan lebih tepat dengan cara ini juga lebih mudah untuk membandingkan performa para siswa yang berbeda dan untuk memperoleh tingkat reliabilitas yang lebih besar dalam penilaian. Jika tugas ‘menulis’ tidak terkontrol, para siswa yang diuji mungkin juga bisa menutupi kelemahan-kelamahnya dengan menjauhi masalah.

Ada beberapa jenis stimuli yang berbeda yang bisa digunakan pada tugas-tugas menulis yang terkontrol. Stimuli bisa ditulis, diucapkan atau non-verbal secara paling efektif, contohnya grafik, rencana atau gambar di mana siswa diminta untuk menjelaskan dengan tulisan. (lihat appendix V; Dunlop, 1969; McEldowney, 1974, 1976, 1982; Weir, 1983a, dan Appendix I untuk contoh-contoh tersebut).

Manfaat

1. Manfaat stimuli non-verbal yaitu jika mereka memberikan informasi dengan cara yang jelas dan tepat, peserta tidak harus menghabiskan banyak waktu dalam memberi sandi suatu teks tertulis. Tugas menjadi yang paling efektif ketika peserta diminta untuk mengomentari kecenderungan khusus yang ditunjukkan dalam sebuah grafik, atau untuk membandingkan dan membedakan satu set gambar dengan gambar lainnya. Stimuli yang berbeda bisa digunakan untuk memperoleh performa tertulis dari beberapa fungsi bahasa yang berbeda seperti argumentasi, deskripsi suatu proses, perbandingan dan perbedaan atau atau tulisan satu instruksi.

Kerugian

1. Permasalahan muncul ketika keinginan untuk tidak menyukai kelompok peserta tertentu manapun. Test diambil bagi daerah yang benar-benar spesial seperti jilid buku, atau pertengahan helm untuk stimulus visualnya. Para peserta sering tidak mengatasi tantangan mental untuk mengambil jenis ini dan menyerah daripada melompati lewat intelektualnya untuk mendapat tugas menulis.

Masalah mungkin selalu terjadi ketika kerumitan stimulus menghalangi hasil yang diinginkan, yakni seseorang harus memahami instruksi-instruksi yang

sangat rumit dan/atau stimuli visual untuk membuat deskripsi yang relatif terus terang mengenai proses atau klasifikasi data.

2. Kesulitan-kesulitan yang disebabkan oleh tugas-tugas tipe transfer informasi ini mungkin muncul lewat perbedaan-perbedaan yang berhubungan dengan pendidikan dan kebudayaan dalam kemampuan untuk menjelaskan grafik atau tabel atau gambar bergaris.

Ringkasan

Manfaat

1. Ringkasan bisa menjadi test yang valid, contohnya sangat tepat untuk menguji kemampuan menulis siswa dengan tugas ini. Siswa harus mengevaluasinya dalam situasi akademis. Menulis laporan atau essay membutuhkan kemampuan untuk memiliki fakta yang relevan dari banyaknya data dan untuk mengkombinasi ulang ini dalam bentuk yang dapat diterima. *Summary* yang melibatkan kemampuan untuk menulis komposisi yang terkontrol yang mengandung ide esensial 'menulis' dan membuang yang non-esensial.

Kerugian

1. Masalah dari kekhususan dari teks para peserta diharapkan membuat tulisan dalam tugas ringkasan seperti dalam tugas-tugas menulis yang terkontrol lainnya. sering dapat kesulitan dalam memilih teks stimulus yang tepat karena kekhususan subjek mereka akan menciptakan beberapa masalah bagi yang bukan ahli dan test mungkin menjadi tidak valid. Satu alternatif yaitu memilih teks yang tidak dikenal yang tidak seorangpun yang menemukan kemampuan pokok.

Jika para siswa jurusan IPA dan teknik mesin harus membaca teks 'umum' atau teks 'netral' dan kemudian merangkum dengan menggunakan kosa kata non-sains dan menunjukkan kualitas tulisan dan imajinasinya mungkin terdapat beberapa masalah validitas yang serius bagi para siswa ini. Meskipun

siswa IPA mungkin tidak bisa merangkum sebagian dan ‘mengapa kucing mungkin menjadi binatang peliharaan yang cocok untuk seorang wanita tua’ ia mungkin bisa merangkum keistimewaan proses yang penting.

2. Kesulitan utama pada komponen menulis yang digabungkan dari jenis ini yaitu membuat penilaian yang dapat dipercaya dan konsisten, menilai jawaban siswa dengan dapat dipercaya seseorang harus merumuskan poin-poin pokok yang masuk ke dalam kutipan, merencanakan skema penilaian yang tepat dan menstandarisasi penilaian.

Manfaat pendekatan impresionik dan analitik terhadap penilaian untuk memperbaiki reliabilitas dan validitas sub-test ‘menulis’ akan diujikan di bawah. Perhatian sedikit diberikan kepada perbaikan reliabilitas penilaian menulis dan satu upaya telah dibuat untuk meneliti suatu bidang dan membawa bersama-sama apa yang diketahui tentang pendekatan-pendekatan pokok terhadap masalah ini. Mengenai seluruh struktur buku ini, menempati sebagian besar pembahasan metode test memberikan pentingnya keterampilan khusus yang krusial kepada para siswa yang belajar lewat medium bahasa Inggris, perlakuan yang ada dianggap ber*Manfaat*. Komentar-komentar bagi penilaian menulis menggunakan, *mutatis muntandis*, bagi penilaian hasil yang dibicarakan. Kita mempunyai hasil yang dapat diidentifikasi yang bisa dievaluasi yang berkenaan dengan kriteria yang spesifik.

4.3.3 Penilaian Kesan Umum dan Analitik

Perbandingan dua pendekatan

Kita telah membahas bagai mana mungkin memperbaiki validitas dan reliabilitas. Kita membuat kesimpulan bahwa ada satu syarat bagi sub-test menulis yang terkontrol yaitu *register*, konteks dan jangkauan tugas menulis telah ditentukan untuk peserta. Pada bagian ini, kita menguji bagaimana aplikasi pendekatan-pendekatan yang berpengaruh dan pendekatan-pendekatan nalisis yang distandarisasi terhadap penilaian mungkin dapat membantu kita dalam upaya untuk memperbaiki reliabilitas dan validitas sub-test menulis.

Penilaian analitis menunjuk kepada metode untuk tiap kriteria yang terpisah pada skema nilai diserahkan pada nilai yang terpisah dan nilai akhirnya merupakan gabungan dari penilaian-penilaian individu itu.

Metode kesan pada penilaian biasanya memerlukan dua penilai atau lebih yang memberi satu nilai berdasarkan total kesan mereka terhadap komposisi sebagai keseluruhan (lihat Luiseman, 1949; E. Ingram, 1970). Tiap kertas diberi skor dengan menggunakan skala yang disetujui dan skor seorang peserta ujian kira-kira dari nilai-nilai gabungan. Gagasan penilaian esan secara spesifik meniadakan beberapa usaha untuk memisahkan ciri-ciri komposisi yang berbeda untuk tujuan-tujuan penilai. Menurut Francis (1977), dalam bentuknya yang paling murni, penilaian kesan biasanya mengharuskan setiap penilai untuk membaca contoh skrip, mungkin 10-255, untuk menentukan standar dalam pikirannya dan kemudian membaca semua skrip dengan cepat dan memberi angka nilai pada setiap skrip.

Hartog dkk (1936) mengadakan satu dari beberapa studi yang ada sekarang ini kedalam keefektifan penilaian kesan analitis dan umum untuk menilai komposisi bahasa Inggris. Ada beberapa maksud untuk mengetahui metode mana yang menghasilkan hasil-hasil yang lebih besar yang berkenaan dengan kemampuan untuk mengurangi kesalahan penilai. Penelitian ini menemukan (h.123) bahwa perbedaan antara para peneliti telah dikurangi oleh metode analitis; 'ada beberapa ketidaksesuaian yang lebih besar antara nilai-nilai yang diketahui oleh kesan daripada antara nilai yang diketahui oleh detail terlihat bahwa ketidaksesuaian ini semestinya pada perbedaan-perbedaan yang lebih besar dari standar-standar penilaian para penguji yang berbeda ketika mereka menilai dengan menggunakan kesan'.

Penelitian juga menunjukkan bahwa sebagian besar penguji menunjukkan konsisten berat sebelah pada hal yang berkenaan dengan kemurahan hati dan kesederhanaan dalam penilaian mereka. Bukti yang mereka dapatkan dari ketidaksesuaian pada urutan penempatan dengan menggunakan beberapa cara lebih serius sejak pertentangan jenis ini tidak mudah bagi koreksi dengan cara yang sama seperti perbedaan yang berasal dari bias susunan nilai. Keduanya bisa

dikoreksi oleh ketentuan skema nilai yang diperinci dan oleh standarisasi para penguji yang efisien berdasarkan teori untuk menilai tugas.

Seperti halnya Hartog dkk (1936), Cast (1939) juga menemukan metode analisis yang agak hebat dalam sistem penilaian tunggal. Kecamannya terhadap metode kesan yaitu bahwa menilai mereka dengan karakteristik yang lebih dangkal daripada metode analisis. Bagaimanapun juga, meskipun metode analisis mempertimbangkan yang lebih sesuai, Cast merasa bahwa hasil-hasilnya tidak memberikan bukti yang pasti dari reliabilitas penilaian analitis yang unggul dan oleh karena itu ia menolak penggunaan salah satu metode itu sendiri.

Cast menunjukkan karakter-karakter penting yang melekat dalam dua sistem. Ciri metode analisis yang penting (h.263-4) adalah: 'dalam merata-ratakan nilai untuk semua pertanyaan, susunannya pasti menyusut... "regresi" ini menjadi konsekuensi yang tidak dapat dihindari dari semua bentuk penyajian terakhir pada perbandingan gambar-gambar yang tidak digabungkan dengan sempurna. Dia mencatat (h.263) bahwa penilaian kesan ditentukan oleh peserta individu dan bahwa angka nilai diketahui oleh penguji yang berbeda pada skrip yang sama yang cenderung luas tidak biasanya.

Cast (h.264) juga mencatat kecenderungan penilaian pesan:

Untuk mengukur beberapa poin yang penting atau dangkal-kesalahan-kesalahan ejaan, tata bahasa atau fakta dan menyusun semua bagian: sebaliknya, metode analitis dengan menggunakan poin-poin terbatas yang banyak dan poin-poin yang tidak esensial yang memungkinkan, mungkin mengabaikan kualitas-kualitas tertentu yang menggolongkan essay sebagai keseluruhan.

Francis (1977) juga menunjukkan bahwa bahaya besar dari penilaian pesan yaitu bahwa pesan kualitas sebagai keseluruhan akan dipengaruhi oleh hanya satu aspek dari aspek-aspek kerja. Dia berpendapat bahwa prasangka dan bias dari penilaian mungkin menjadi bagian yang lebih besar dalam menentukan nilai daripada dalam skema analitis.

Penilaian Multiple (ganda)

Wiswman (1949) meneliti kemungkinan untuk memperbaiki penilaian dengan menjumlah nilai-nilai ganda dari empat penilai bebas, penilai yang tidak distandarisasi, menggunakan metode kesan yang cepat. Dia menemukan bahwa penilaian ganda dengan menggunakan metode kesan telah memperbaiki reliabilitas dan lebih cepat daripada prosedur-prosedur analitis perbandingan. Dia (h.205) memperkirakan bahwa jika rata-rata inter-korelasi sebuah kelompok dari empat penilai lebih rendah 0,6 dari yang lainnya: 'perkiraan korelasi yang mungkin dari nilai-nilai rata-rata 'betul adalah 0,92. Nilai ini sangat tinggi dibanding dengan yang kami perkirakan dari satu nilai analitik'.

Wiseman (h.208) mengambil contoh kesakitan menjadi stres bahwa: 'efisiensi penilai harus menilai atas dasar konsistensinya.' Dia berpendapat (h.2204) bahwa konsistensi efisien diperoleh dengan nilai yang murni, penilaian korelasi (penilai-intra reliabilitas), menggunakan metode penilai yang sama pada kedua kesempatan: 'merupakan satu ukuran yang jelas atas konsistensi yang benar, dan satu yang menutup hubungan kepada konsep normal dari reliabilitas test.' Dengan menggunakan sistem penilaian ganda yang berdasar pada prinsip konsistensi diri sendiri, dia memungkinkan untuk mencapai level reliabilitas yang tinggi.

Karya Coffman dan Kurfman (1968) dan Wood dan Wilson (1974) menggambarkan peringatan bagi permasalahan instabilitas penguji dengan menilai perilaku. Mereka telah membuat fakta bahwa menilai perilaku tidak lagi seimbang selama masa penilaian, ketika nilai yang besar dari sebuah skrip dinilai (lihat Edgewort, 1888). Mereka berargumen atas subjektifitas setiap skrip yang lebih dari satu penilai, yang mungkin membantu menetralisasi efek dari inkonsistensi penilaianm perilaku selama penilaian berlangsung.

Walau beberapa keraguan diekspresikan pada masa lampau (lihat Edgewort, 1888) mengenai kelayakan memiliki lebih dari penilai, Britton (1963), Britton dkk (1966), Lucas (1971) dan Wood Quinn (1976) semuanya menemukan bahwa penilaian ganda memperbaiki reliabilitas penilaian essay bahasa Inggris.

Britton dkk (1966), dalam sebuah eksperimen yang dirancang untuk menemukan reliabel yang lain yang menilai perlengkapan untuk digunakan oleh

dewan ujian, membandingkan penilaian ganda ekperimental dengan penilaian tunggal yang diangkat oleh dewan uji GCE. Mereka menemukan (h.21): ‘figur-figur itu jelas mengindikasikan bahwa dalam kasus penilaian oleh penguji secara individu dengan uraian yang sangat hati-hati dan meneliti aransemen bagi modernisasi pada kenyataannya kurang reliable dibanding penilaian ganda. ‘ketika ofisial menilai dan penilaian ganda dikorelasikan dengan kriteria eksternal dari pekerjaan rumah yang dibuat oleh para peserta selama tahun ajaran, penilaian ganda ditmukan untuk mencocokkan.

Head (1966) membuat suatu percobaan untuk menemukan tambahan penilaian impresi daru dua penguji akan lebih reliabel dibanding dengan penguji individual. Dia menemukan (h.71): ‘kenaikan koefisien dari 0,64 bagi korelasi penilaian tunggal menjadi 0,84 untuk korelasi pasangan penilaian yang menunjukkan bahwa penilaian tambahan menjadi lebih reliabel.’

Lucas (1971) menmukan bahwa meskipun penggunaan penilai yang inkonsisten (artinya nilai korelasi hanya 0,65) dalam penilaian ganda dengan impresi menambah reliabilitas pemberian nilai yang signifikan. Pertambahan yang besar dalam reliabilitas terjadi dalam perubahan dari satu menjadi dua peniali.

Wood dan Quinn (1976) menggunakan level ‘O’ essay bahasa Inggris dan pertanyaan *summary* menemukan bahwa penilaian impresi oleh sepasang penilai menjadi lebih reliabel dibanding dengan penilai tunggal. Mereka memberi kesan, walaupun, tidak terdapat suatu keuntungan dalam reliabilitas dari suatu penilaian analitik dibanding dengan penilaian impresi tunggal. Pembaharuan yang ril adalah dalam hal penilaian.

Penilaian Holistik

Dalam evaluasi holistik, penilai berdasar kepada penilaian mereka dalam impresi mereka dari semua komposisi: dalam penilaian frekuensi (lihat Steel dan Talman, 1936), total penilai atau menjumlahkan berbagai elemen dalam komposisis, seperti: perlengkapan kohesif, kesalahan ucap kata, kesalahan peletakan koma, atau kesalahan kalimat. Jacobs dkk berpendapat bahwa metode ayang akan datang lebih objektif dan juga lebih reliabel. Validitasnya tidak begitu

pasti karena satu komposisi dievaluasi oleh metode *frequency-count* yang telah dinilai bukan untuk efek komunikatif, tetapi untuk nomor atau salah satu dari elemen.

Evaluasi holistik terlihat lebih objektif dibanding impresi yang dibuat oleh para penilai. Jacobs dkk (h.29) berpendapat bahwa:

Atas kedengkian (atau malah karena) subjektifitas ini, evaluasi holistik telah memperlihatkan kapabilitas dalam membuat penilaian reliabel yang tinggi. Kebanyakan para pelajar menguji... pada faktanya, berdasar pada evaluasi holistik dari satu tipe atau yang lain dan semua pelajar memperoleh reliabilitas pembaca dalam tengah-tengah-menuji-tinggi delapan puluh atau sembilan puluhan. Secara intuitif hal itu terlihat bahwa komposisi skor berdasar pada respon holistik dari para pembaca yang menyertai pesan penulis harus lebih valid daripada yang berdasar pada metode *frequency-count*, yang pada pembayaran terbaik hanya sebuah kecupan bagi pendapat dan ide penulis. Sebagaiman Cooper (1977) menaruh hal itu, 'evaluasi holistik oleh responden manusia menjadikan kita lebih dekat kepada sesuatu yang esensial manusia komunikasi dibanding apa yang dilakukan oleh *frequency-count*.'

Evaluasi holistik jelas dlebih-lebihkan di man perhatian utama adalah dengan mengevaluasi ketidak-efektifan komunikatif tulisan peserta. Itu adalah sebuah kasus dalam proyek TEEP (lihat Weir, 1983a, dan Appendix I) di mana preferensi diperuntukkan bagi analitik, skema penilaian holistik dalam impresionisti yang satu, menyokong sesuatu yang eksplisit daripada daftar yangimplisit dari sautu roman atau kualitas untuk membantu para penilai.

Hal itu terasa sangat kuat, bahwa perhatian yang sangat kecil telah terbayar pada masa lalu dengan kriteria aktual yang diaplikasikan, secara implisit atau eksplisit, menjadi contoh dari pembuatan karya tulis. Sejajar dengan skema anlaitik yang diserahkan kepada para pelajar, terlalu banyak ruang bagi interpretasi idiosinkratik dari segala standar konstitut yang masih diaplikasikan kedalam skrip. Aplikasi yang bersih, kriteria yang tepat telah dirasa begitu penting.

Jacobs dkk membuat perbedaan anatara penilain holistik dan penilaian jumlah frekuensi seperti terhadap divisi yang saling melengkapi kedalam penilaian kesan dan penilaian analitis yang digunakan oleh para peneliti. Mereka

menggambarkan divisi sebagai berikut: “istilah holistik Cooper berarti beberapa prosedur yang berarti “mengurutkan ciri-ciri retorik dan ciri-ciri informasional.”

Chaplen (1970an) berpendapat bahwa hasil-hasil yang lebih reliabel mungkin bisa diperoleh dari metode kesan untuk menilai apakah skala yang digunakan merupakan salah satu yang tiap nilainya disamakan dengan tingkat penerimaan yang berbeda. Ini adalah pendekatan yang digunakan oleh British Council dalam sistem ujian ELTS. Itu mungkin digambarkan sebagai suatu kesan berdasarkan sistem pemberian tanda. Contoh dari skema nilai yang ditandai ini bisa ditemukan dalam buku karya B.J Carroll (1980b, hal.136).

Pendekatan carroll bagus dalam konsep seperti memberikan deskripsi yang lebih detail untuk institusi. Masalahnya ia gagal dalam praktek karena tidak melayani para pelajar yang tingkat performa berbeda dipandang dari segi kriteria yang berbeda. Seorang kandidat mungkin diberi nilai 7 untuk ‘kelancaran’, dan nilai 5 untuk ‘keakuratan.

Masalah gagalnya kriteria ini dihindari oleh skema nilai yang lebih ‘analitis’, untuk level-level tiap kriteria dan untuk suatu pengukuran yang paling integratif. Metode ini memiliki manfaat lain yang akan menjadikannya lebih mudah untuk menyelesaikan laporan profil dan bisa menunjukkan peran diagnostik dalam menggambarkan kelebihan dan kelemahan hasil tulisan siswa.

Skema nilai analitis tampak seperti suatu alat yang jauh lebih berguna untuk melatih dan menstandarisasi para penguji yang baru. Fancis(1977) menjelaskan bahwa dengan menggunakan skema analitik, menguji isi bisa lebih baik dengan melatih dan menstandarisasi para penguji yang baru tentang kriteria penilaian. Ukuran persetujuan tentang apa yang tiap kriteria harapkan bisa ditentukan, dan para penilai bisa distandarisasi terhadap level yang berbeda dari tiap kriteria ini.

Skema nilai analitis ditemukan dari upaya untuk membuat penilaian lebih objektif. Brooks (1980) menunjukkan bahwa kualitas yang dinilai dengan menggunakan skema nilai analitis di masa lalu sering sulit untuk dipahami. Maka, meskipun skema analitis mungkin memfasilitasi persetujuan diantara para penguji, subjektivitas yang terlibat dalam penilaian pada beberapa skema mungkin sedikit

direduksi karena kurangnya ketegasan terhadap kriteria yang dapat dipakai, atau melalui penggunaan yang tidak jelas.

Menentukan Kriteria yang Tepat untuk Menilai Hasil Tertulis:

Test Bahasa Inggris untuk Tujuan-tujuan Pendidikan (TEEP)

Gagalnya skema analitis di masa lampau telah menjadi pilihan dan gambaran dari kriteria yang tepat bagi situasi yang diberikan. Dalam model test TEEP (lihat Weir, 1983a dan Appendix I) terasa bahwa penilaian contoh-contoh performa tertulis berdasarkan pada kriteria analitis yang tepat yang digolongkan berdasarkan level-level performa yang berbeda.

Data yang melaporkan seleksi kriteria penilaian datang dari survey yang dilakukan terhadap para guru bahasa di sekolah ARELS. Bukti empiris dikumpulkan dari 560 dosen untuk membantu memutuskan kriteria yang bisa digunakan untuk menilai jenis latihan-latihan transfer informasi tertulis yang terjadi dalam konteks akademik.

Sebagai hasil dari investigasi kriteria korelevanan dan cukup, komposisional, organisasi, kohesi, cukupnya referensi, keakuratan gramatikal, ejaan dan tanda baca terlihat seperti yang paling cocok untuk menilai tugas-tugas menulis.

Untuk menggunakan kriteria yang 'valid', satu upaya telah dilakukan untuk membuat skema penilaian analitis yang tiap kriteriannya dibagi menjadi empat level behavioral pada skala 0-3 (lihat tabel di bawah). Level 3 dapat disamakan dengan kompetensi minimal. Dalam level ini terasa bahwa siswa mungkin mempunyai sedikit masalah yang berhubungan dengan tugas-tugas menulis. Pada level 2, beberapa masalah yang berhubungan dengan kriteria ini muncul, dan bantuan remedial sebaiknya dilakukan. Level 1 akan mengindikasikan bahwa banyak bantuan yang perlu dilakukan terhadap kriteria ini. Level 0 menunjukkan hampir semua tidak kompeten dalam respect terhadap pertanyaan.

Skala-skala Penulis Atribut TEEP

A. Relevansi dan Cukupnya Isi

0 - Jawaban hampir tidak mengandung isi yang berhubungan dengan tugas.

Total jawabannya tidak cukup.

1- Jawaban mempunyai relevansi yang terbatas terhadap tugas.

Memungkinkn adanya celah-celah besar dalam melaporkan topik dan/atau adanya pengulangan yang tidak berarti.

2- Sebagian besar tugas terjawab, meskipun mungkin ada beberapa celah atau informasi yang berlebih-lebihan.

3 - Jawaban sudah relevan dan cukup.

B. Organisasi Komposisional

0 - Tidak adanya organisasi isi yang jelas.

1- Organisasi isi yang ada hanya sedikit.

2- Terdapat beberapa keterampilan organisasional;, tapi belum terkontrol dengan baik.

3- Penggunaan kohesi dan pola internal sudah cukup jelas, keterampilan-keterampilan organisasional cukup terkontrol

C. Kohesi

0- Hampir tidak ada kohesi. Tulisan sangat tidak lengkap dan secara virtual tidak mungkin adanya pemahaman terhadap komunikasi yang dimaksud.

1- Kohesi yang tidak memuaskan mungkin menyebabkan sulitnya memahami sebagian besar komunikasi yang dimaksud.

2- Sebagian besar kohesi memuaskan mekipun kadang-kadang kurang memuaskan, mungkin berarti bahwa bagian-bagian tertentu dari komunikasi tidak selalu efektif.

3- Penggunaan kohesi yang memuaskan menghasilkan komunikasi yang efektif.

D. Cukupnya Kosakata

0- kosakata yang dimiliki tidak cukupnya bahkan untuk komunikasi yang paling dasar.

1- Kosakata yang dimiliki untuk tugas tidak cukup. Mungkin leksikalnya sering tidak tepat dan/atau sering ada pengulangan-pengulangan.

- 2- Ada beberapa kosakata yang tidak cukup untuk tugas. Mungkin ada beberapa leksikal tidak tepat dan/atau adanya pemakaian kata yang berlebih-lebihan.
- 3- Hampir memiliki cukup kosakata untuk tugas. Hanya kadang-kadang tidak cocok dan/atau berlebihan.

E. Grammar

- 0- Semua pola gramatikal sering tidak akurat.
- 1- Grammar sering tidak akurat.
- 2- Beberapa grammar tidak akurat.
- 3- Hampir tidak ada grammar yang tidak akurat.

F. Tanda Baca

- 0- Ketidaktahuan konvensi tanda baca.
- 1- Standar rendah dari keakuratan tanda baca.
- 2- Beberapa tanda baca tidak akurat.
- 3- Hampir tidak ada tanda baca yang tidak akurat.

G. Ejaan

- 0- Hampir semua ejaan tidak akurat.
- 1- Standar rendah terhadap keakuratan ejaan.
- 2- Beberapa ejaan tidak akurat.
- 3- Hampir tidak ada ejaan yang tidak akurat.

Pertimbangan Berikutnya dalam Membuat Pola Tugas-tugas menulis untuk Dimasukkan ke dalam Rangkaian Test

Nomor Tugas-tugas Menulis

Pembahasan menulis yaitu mengenai bagaimana reliabilitas penilai dapat tercapai. Ada faktor-faktor lain yang memiliki kontribusi pada reliabilitas test. Pertama, beberapa contoh dari karya siswa yang diambil bisa membantu mengontrol perbedaan performa yang mungkin terjadi dari tugas ke tugas.

Baik reliabilitas maupun validitas yang ditingkatkan dengan cara penarikan contoh lebih banyak daripada dengan satu komposisi dari tiap kandidat. Finlayson (1951, hal. 132) melihat bahwa “performa dari seorang anak

pada satu essay tidak representatif terhadap kemampuannya menulis essay secara umum”. Penelitian Vernon dan Milligan (1954, hal.69) juga memperoleh bahwa “ada keraguan yang sangat besar terhadap praktek biasa...mencoba untuk menilai kemampuan bahasa Inggris umum dari satu essay yang dinilai oleh seorang penguji”.

Ebel (1972) menunjukkan bahwa lebih banyak contoh dari tulisan siswa pada suatu test, maka hasilnya akan lebih reliabel. Ebel menguraikan bagaimana skor test terdiri dari dua unsur: skor yang benar dan skor yang salah.

Murphy (1978) juga berpendapat bahwa faktor penting dalam menentukan reliabilitas yang bermacam-macam dari delapan ujian GCE di bawah ini:

Beberapa nilai untuk individu yang berkontribusi dalam nilai-nilai ujian akhir. Efek peningkatan reliabilitas yang dilakukan dengan cara memiliki lebih banyak bagian dari ujian ditunjukkan oleh kasusu bahasa Inggris level “A”. Observasi ini konsisten terhadap prinsip yang telah ditentukan yang mana kombinasi ukuran-ukuran yang tidak reliabel menjadi lebih reliabel dari ukuran-ukuran individu itu sendiri.

Jacobs dkk (1981, hal.15) berpendapat bahwa:

Sebaiknya memperoleh paling tidak dua komposisi dari tiap siswa. Bantuan-bantuan ini memastikan bahwa test tersebut melakukan penarikan contoh representatif dari kemampuan penulis, dengan mereduksi beberapa efek dari variasi performa individu dari topik ke topik atau dari satu periode test yang lain... pengalaman kita yang lainnya mengaharapkan dua tugas menulis yang dirumuskan dengan hati-hati cukup memungkinkan bagi kebanyakan situasi ujian.

Nyatanya lebih banyak contoh tulisan siswa yang diambil lebih baik dari ini akan menjadi tujuan reliabilitas dan validitas, menjadikan tiap contoh memperkirakan kemampuan yang bisa dipertanggungjawabkan.

Pilihan Pertanyaan

Seperti seleksi topik, penting untuk memastikan bahwa siswa mampu menulis sesuatu dengan topik yang diberikan. Apakah berarti membiarkan pemilihan topik itu merupakan suatu keputusan yang penting yang harus dilakukan, untuk itu bisa mempengaruhi reliabilitas test.

Jacobs dkk (1981, hal.1) menyatakan:

Untuk evaluasi skala besar, sebaiknya semua siswa menulis topik yang sama karena dengan membiarkan memilih topik akan memasukkan begitu banyak perbedaan yang tidak terkontrol kedalam test, artinya apakah perbedaan skor yang diobservasi harus menjadi perbedaan-perbedaan nyata dalam kecakapan menulis atau topik-topik yang berbeda? Tidak ada dasar reliabel bagi perbandingan skor pada suatu test jika semua siswa tidak mengerjakan tugas menulis yang sama; selain itu, konsistensi atau reliabilitas pembaca dalam mengevaluasi test mungkin akan direduksi jika semua bacaan dari sesi penskoran tunggal bukan merupakan topik yang sama.

Heaton (1975) berpendapat bahwa menyediakan pilihan berarti siswa akan menghabiskan banyak waktu untuk mencoba memilih topik dari beberapa alternatif yang diberikan. Dimana test-test yang dilakukan dengan waktu yang terbatas, memaksa siswa untuk menulis dengan topik yang sama mungkin juga bermanfaat bagi kandidat yang tidak jelas. Jacobs dkk (1981, hal.17) menyimpulkan:

Mengingat masalah-masalah yang berhubungan dengan penyediaan pilihan topik, alternatif terbaiknya, jika keterampilan memilih topik bukan diantara tujuan-tujuan test, akan terlihat mengharuskan semua siswa untuk menulis dengan topik yang sama, dan untuk memberi mereka lebih dari satu kesempatan untuk menulis.

Dengan mendasarkan tugas-tugas menulis pada teks tertulis atau lisan yang diberikan kepada kandidat atau stimuli non-verbal, mungkin untuk memastikan bahwa pengetahuan subjek semuanya mulai sama paling tidak yang berkenaan dengan informasi yang tersedia untuk mereka. Semuanya dibutuhkan untuk menulis dengan topik yang sama, tapi mereka akan menulis dengan topik yang berbeda-beda.

Waktu yang Diberikan bagi Tiap Tugas Menulis: Percabangan Limit Waktu

Jacobs dkk (1981, hal.17) menunjukkan perlunya memberi pertimbangan kepada tujuan test menulis:

Apakah test hasil perkembangan langsung dari aktivitas belajar tertentu, mungkin termasuk, revarasi untuk komposisi test (embaca buku-buku tertentu atau melakukan penelitian topik yang ditentukan, mempraktekan

topik yang sama tau mode yang sama di dalam kelas), atau apakah test dadakan, yang memusatkan pada hasil gubahan, daripada proses gubahan?

Jacobs dkk (1981, hal.17) menunjukkan beberapa cabang dari perbedaan ini:

Test dadakan dengan waktu terbatas bisa mulai dengan memberi sumber-sumber penulis kepada semua proses menulis, dengan hasil yang mirip dengan apa yang biasa penulis lakukan pada proses menulis. Penting untuk mengingat limitasi waktu.

Waktu yang tepat bagi penyesuaian tugas-tugas menulis orientasi hasil dalam ujian biasa. Jacobs dkk (1981, hal.18) berpendapat bahwa:

“Test komposisi yang diberikan bersama dengan rangkaian pengukuran lain harus membatasi waktu jika semua waktu test menjadi praktis dan tidak mengenalkan beberapa perbedaan yang pasti akan membosankan bagi peserta ujian... kita memberi batasan waktu 30 menit untuk test komposisi yang diberikan sebagai bagian dari Test Michigan dan waktu yang diberikan itu cukup untuk menghasilkan contoh kemampuan menulis mereka.dalam penelitian mereka (hal.19), mereka menemukan bahwa”Dengan test komposisi selama 30 menit... tapi sebagian besar siswa dengan kemampuan level dasar umumnya bisa menulis sekitar satu halaman atau lebih.”

5.3.4 Kesimpulan

Komponen menulis dan beberapa test akan pada tugas-tugas menuli yang dikontrol dimana ciri-ciri audiensi, medium, keadaan, dan tujuan bisa lebih spesifik. Perhatian harus diberikan pada perkembangan kriteria penskoran yang cukup dan tepat serta pada para penguji yang dilatih dan standarisasi terhadap penggunaan ni.

5.4 Ujian Speking (Berbicara)

Ujian kemampuan berbicara memberikan cukup kesempatan untuk menemukan kriteria untuk ujian komunikatif, artinya bahwa: tugas-tugas yang dikembangkan dalam paradigma ini akan mempunyai tujuan, akan menarik, dan mempunyai motifasi, dengan efek wasback positif pada pengajaran yang mendahului test; interaksi akan menjadi ciri kunci; akan ada tingkat

intersubjektifitas diantara partisipan; hasilnya akan menjadi tidak bisa diprediksi; konteks realistis akan diberikan; dan pengolahan akan dilakukan. Mungkin lebih banyak dari beberapa keterampilan lain. Ada kemungkinan untuk membnagunnya menjadi test karakteristik dinamis dari komunikasi aktual (lihat bagian 3.1).

Masalah-masalah penialian kemampuan berbicara lebih besar dari penilaian menulis karena interaksinya berlalu dengan cepat dan tidak bisa dicek. Tugas penting bagi pembuat model test harus menentukan aktifitas apa yang harus kandidat tunjukkan, seberapa jauh karakteristik komunikatif dinamis yang berhubungan dengan aktifitas-aktifitas ini bisa dimasukkan kedalam test, dan dimensi tugas apa yang akan melibatkan kompleksitas, ukuran susunan percakapan referensial dan fungsional untuk diproses dan dihasilkan.

5.4.1 Essay Verbal

Kandidat diminta untuk berbicara selama tiga menit dengan satu topik umum yang ditentukan atau lebih.

Manfaat

1. Kandidat harus berbicara panjang lebar yang memungkinkan untuk menggunakan kriteria-kriteria termasuk kefasihan. Pertanyaan-pertanyaan singkat yang berbeda yang harus siswa jawab sering membatasi susunan kriteria yang dapat diaplikasi.

Kerugian

1. Masalah-masalah yang berhubungan dengan tugas menulis tidak terkontrol bebas menggunakan jenis lisan ini. Topik ditentukan mungkin tidak menarik bagi kandidat dan bukan sesuatu yang meminta kita untuk melakukannya dalam nyata tanpa persiapan.
2. Lebih banyak open-ended topik, maka performa yang ada lebih sukses tergantung pada pengetahuan latar belakang dan pengetahuan kultural dan menggunakan faktor-faktor seperti imajinasi atau kreatifitas. Devinisi

respon-respon terhadap apa yang diharapkan dari isi lebih sulit untuk mempertahankan reliabilitas dalam penilaian.

3. Penggunaan tape recorder dalam tugas ini mungkin menjadi tekanan bagi para kandidat.

5.4.2 Presentasi Lisan

Kandidat diharapkan untuk berbicara singkat dengan topik yang telah dia siapkan sebelumnya. Berbeda dari “Essay Lisan”.

Manfaat

1. Sangat efektif untuk membuat kandidat menceritakan dirinya sendiri. Dalam test TEEP ini diharapkan sebagai latihan, tapi diketahui bahwa satu menit yang diberikan kepada kandidat untuk berbicara tentang kehidupan pribadinya memberikan semua indikator yang baik dari kecakapan bahasa lisannya yang berkenaan dengan kriteria yang digunakan dalam menilai semua tugas lain. Apa yang penting dalam menilai hasil berbicara memperoleh contoh sufisien dari ucapan kandidat bagi penilaian yang pantas.
2. Mengintegrasikan aktifitas dengan mendengarkan atau membaca teks tugas lisan bisa dicocokkan dengan tugas kehidupan nyata yang kandidatnya harus perform dalam situasi target.

Kerugian

1. Jika kandidat mengetahui topik dengan baik sebelumnya, dia bisa mempelajarinya dengan baik. Jika waktu yang diberikan untuk persiapan sedikit kemudian dia menghadapi masalah yang akan diuji mungkin pengetahuan bukan sebagai kemampuan linguistik. Jika tugas dihubungkan dengan membaca berdasarkan teori untuk memastikan bahwa semua kandidat memiliki informasi yang biasa kemudian dia dihadapkan dengan masalah membaca yang mungkin mengganggu nilai.
2. Keragaman interpretasi mungkin akan menimbulkan masalah dalam penilaian.

5.4.3 Wawancara Bebas

Jenis wawancara ini yaitu percakapan mengembangkan model yang tidak berstruktur dan tidak ada prosedur-prosedur yang ditentukan sebelumnya.

Manfaat

1. Karena validitas permukaan dan isinya, wawancara merupakan alat untuk menguji keterampilan lisan para kandidat.
2. Wawancara bebas yaitu mirip percakapan yang agak lama dan petunjuk yang diberikan untuk melakukan wawancara. Percakapan mungkin terlihat lebih teliti terhadap pola normal dari interaksi sosial yang tidak formal dalam kehidupan nyata dimana tidak ada agenda yang dirumuskan dengan jelas.

Kerugian

1. Karena tidak ada prosedur-prosedur untuk memperoleh bahasa, performa-performa mungkin berbeda dari satu peristiwa keperistiwa lain karna topik-topik yang berbeda mungkin mulai dibicarakan dan perbedaan-perbedaan ini terjadi dengan wawancara.
2. Prosedur ini memerlukan waktu yang banyak dan sulit untuk dilaksanakan jika ada banyak kandidat.

5.4.4 Wawancara Terkontrol

Dalam prosedur ini terdapat prosedur-prosedur yang ditentukan sebelumnya untuk memperoleh performa. Wawancara FSI mirip dengan model ini (lihat Adams dan Frith, 1979 dan Wilds, 1975).

Manfaat

1. Kemungkinan besar para kandidat diberi pertanyaan yang sama dan oleh karena itu lebih mudah untuk membandingkan performa tiap kandidat.
2. Prosedur ini mempunyai tingkat yang lebih tinggi dari validitas isi dan permukaan daripada teknik-teknik lain selain dari latihan-latihan role play

dan celah informasi di UCLES/RSA pada keterampilan komunikatif bahasa Inggris (lihat Appendix III).

3. Dengan latihan dan standarisasi yang cukup dari penguji terhadap prosedur-prosedur dan skala-skala yang digunakan, figur-figur reliabilitas yang dapat dipertanggungjawabkan dapat tercapai dengan menggunakan teknik ini. Clark dan Swinton (1979) melaporkan rata-rata reliabilitas intra-rater 0,867 dan reliabilitas inter-rater 0,75 untuk wawancara jenis FSI.
4. Wawancara lisan yang efektif bisa terjadi ketika kandidat diwawancara dan dinilai oleh ahli bahasa dan ahli subjek yang telah distandarisasi.

Kerugian

1. Salah satu kekurangan wawancara yaitu bahwa ia tidak bisa mencakup kandidat-kandidat yang mungkin mendapatkan dirinya sendirinya terlibat di dalamnya bahkan dimana performa-performa level target yang terbatas seperti pada kasus FSI. Dalam wawancara, sulit untuk meniru semua sifat komunikasi kehidupan nyata seperti timbal balik, motivasi, tujuan dan ketetapan.
2. Bahkan ketika prosedur-prosedur untuk memperoleh performa yang ditentukan sebelumnya masih tidak ada jaminan dimana para kandidat akan diberi pertanyaan yang sama dengan cara yang sama bahkan dengan penguji yang sama.

5.4.5 Transfer Informasi: Deskripsi Urutan Gambar

Kandidat harus memahami panel gambar-gambar yang menggambarkan rangkaian peristiwa secara kronologis dan harus menceritakan kisah pada masa lalu berdasarkan gambar. Sebelumnya waktu diberikan kepada kandidat untuk mempelajari gambar-gambar tersebut.

Manfaat

1. Tugas yang diperlakukan para kandidat sudah jelas. Tidak mengahruskan mereka untuk membaca dan mendengarkan dan oleh karena itu menghindari

kritik kontaminasi penilaian yang memberikan gambar-gambar yang berat sebelah secara kultural atau edukasional.

2. Metode ini akan menjadi prosedur yang efisien dan salah satu dari yang ada untuk membuat kandidat memberikan contoh percakapan yang berhubungan dan membiarkan aplikasi kriteria dalam penilaian. Juga berguna untuk memperoleh kemampuan kandidat untuk menggunakan bentuk-bentuk gramatikal seperti bentuk past tense untuk kalimat tidak langsung (melaporkan).
3. Karena semua kandidat didesak oleh informasi yang diberikan oleh gambar-gambar yang memberikan perbandingan terhadap kandidat yang tidak dipengaruhi oleh pengetahuan latar belakang dan pengetahuan kultural yang telah diberikan.
4. Nilai teknik tergantung pada gambar-gambar yang jelas dan tidak ambigu dan bebas dari bias kebudayaan dan pendidikan. Teknik ini jelas dan disukai oleh dewan pengurus ujian sekolah British. Dalam studi dengan format yang sesuai untuk komponen lisan bagi TOEFL (Clark dan Swinton, 1979) ini terbukti menjadi satu dari format-format yang paling efektif dalam test eksperimental.

Kerugian

1. Keaslian tugas ini terbatas meskipun bisa dikatakan mewakili situasi yang harus menggambarkan sesuatu yang terjadi yaitu rutinitas informasional. Ini mungkin benar-benar menjadi fungsi yang paling penting dalam beberapa kesempatan. Ia memberi tahu sedikit tentang kemampuan kandidat untuk berinteraksi secara lisan.
2. Jika kualitas gambar tidak sempurna maka kandidat tidak mungkin memiliki kesempatan untuk menunjukkan performa terbaiknya. Perbedaan-perbedaan dalam interpretasi mungkin juga menunjukkan ketidak reliabelan terhadap penilaian.

5.4.6 Transfer Informasi Pertanyaan-pertanyaan dari Satu gambar

Penguji memberikan sejumlah pertanyaan tentang isi gambar yang telah kandidat pelajari. Pertanyaan-pertanyaan mungkin diharapkan mencakup pikiran-pikiran dan sikap orang-orang di dalam gambar kemudian mendiskusikan perkembangan-perkembangan dari apa yang digambarkan.

Manfaat

1. Mungkin ada *Manfaat* yang dipertimbangkan dalam meneliti teknik ini, yang telah menunjukkan peran yang ada dalam komponen lisan dari test bahas Inggris PLAB bagi para dokter di luar negeri. Dalam PLAB para kandidat diberi petunjuk slide, x-rays, gambar-gambar kondisi medis, dan lain-lain, kemudian diminta untuk memberi pendapat tentang petunjuk itu seperti menjawab pertanyaan-pertanyaan yang berhubungan dengan petunjuk tersebut.

Kerugian

1. Kandidat merupakan pelaku satu-satunya dalam peran responden dan menghilangkan kesempatan untuk memberi pertanyaan. Kriteria tibal-balik, ciri normal dari sebagian besar interaksi lisan tidak diketahui.
2. Gambar-gambar harus jelas dan tegas untuk alasan-alasan yang telah diuraikan pada pembahasan urutan gambar-gambar. Jika sebagian besar kandidat harus diuji selama beberapa hari maka pertanyaan tentang test akan muncul jika gambar-gambar yang sama harus digunakan. Tapi jika gambar-gambar berbeda terpaksa digunakan maka persoalan tentang hal yang bisa diperbandingkan harus dihadapinya.

5.4.7 Tugas-tugas interaksi

Celah Informasi Siswa-siswa

Dalam tugas ini biasanya para siswa bekerja berpasangan dan masing-masing hanya diberi bagian informasi yang penting untuk menyelesaikan tugas. Mereka harus menyelesaikan tugas dengan mencari informasi yang hilang dari

yang lain. Para kandidat harus berkomunikasi untuk mengisi celah informasi dalam situasi yang bermakna.

Sertifikat UCLES/RSA untuk keterampilan komunikatif bahasa Inggris memiliki contoh-contoh realistis tertentu dari ini (lihat Appendix III). Seperti perkembangan dari interaksi lawan bicara ini muncul setelah diskusikan dan para kandidat harus melaporkan kesimpulan yang diambil dan memberikan alasan keputusan-keputusan yang diambil.

Manfaat

1. Ada beberapa tugas test yang menggambarkan tindakan komunikasi yang lebih baik untuk memenuhi sebagian besar kriteria yang diberikan oleh Morrow (1979) bagi apa yang menjadikannya komunikatif. Para kandidat akan bebas untuk memilih partner mereka berinteraksi dengan orang-orang yang mereka kenal dan merasa senang berkomunikasi dengan mereka.
2. Seperti ciri interaksi yang normal mereka bisa menggunakan bentuk-bentuk pertanyaan, memperoleh informasi, membuat pertanyaan, meminta klarifikasi dan menguraikannya dengan kata-kata sendiri agar sukses dalam test.
3. Tugas itu sangat interaktif dan menjadi lebih pendek dari sebagian besar tugas yang lain yang menggambarkan komunikasi nyata.

Kerugian

1. Ada satu masalah jika seseorang partisipan mendominasi interaksi dari kandidat lainnya pada kesempatan yang terbatas untuk menunjukkan potensi komunikasinya.
2. Sama halnya dengan jika ada perbedaan besar pada kecakapan antara dua orang mungkin mempengaruhi performa dan pertimbangan yang dibuat.
3. Juga ada masalah jika seseorang kandidat lebih tertarik pada suatu topik atau tugas seperti interaksi yang mungkin menjadi suatu hasil.
4. Performa para kandidat dinilai dalam situasi, dan harus memperkirakan kemampuan performa mereka dalam situasi lain.

5. Juga ada ketidakleluasan praktek seperti waktu yang tersedia, kesulitan-kesulitan administrasi.

Celah Informasi Siswa-Penguji

Untuk menghindari kemungkinan dari ketidakseimbangan pada kontribusi kandidat terhadap interaksi, beberapa dewan pengurus harus mempunyai penguji yang berperan sebagai salah satu partisipan atau menggunakan lawan bicara biasa, seperti guru yang dikenal agar para kandidat merasa nyaman.

Untuk menguji para kandidat secara terpisah mereka bisa diberi diagram, catatan-catatan, dan lain-lain dari informasi yang hilang dan tugas mereka yaitu harus meminta informasi dari penguji.

Manfaat

1. Manfaat utamanya yaitu bahwa ada kesempatan yang lebih besar dimana lawan bicara akan memberi reaksi dengan cara yang sama dengan semua kandidat yang memberikan perbandingan performanya yang lebih pantas.

Kerugian

1. Berinteraksi dengan seorang guru sering menjadi tugas yang lebih menakutkan bagi para kandidat daripada berinteraksi dengan teman sebayanya.
2. Ada beberapa bukti bahwa ketika penguji menjadi seorang partisipan dalam berinteraksi, kadang-kadang dia kurang hati-hati meniali performanya sendiri dan performanya kandidat (Fisher, 1979).

5.4.8 Role-Play

Beberapa dewan pengurus ujian, contohnya AEB dan UCLES/RSA, termasuk role play dimana kandidat diharapkan untuk memainkan salah satu peran dalam interaksi yang mungkin diharapkan dalam dunia nyata. Interaksi bisa terjadi antara dua siswa atau biasanya penguji berperan sebagai salah satunya. *Kerugiannya* yaitu sulit untuk membuat penilaian pada waktu yang

bersamaan ketika dia ambil bagian dalam interaksi. Seperti dalam latihan celah informasi yang melibatkan guru sebagai lawan bicara dan penguji, bahayanya yaitu bahwa nilai yang diberikan akan mewakili gambaran akhir dari performanya sendiri seperti performa siswa.

Manfaat

1. Teknik ini bisa menjadi valid pada persoalan permukaan dan isi untuk situasi-situasi yang bereda dan pengalaman dewan pengurus ujian memberi kesan bahwa itu merupakan prsktek dan secara potensial merupakan alat-alat yang sangat valid dan reliabel untuk menilai kemampuan kandidat dalam berpartisipasi secara efektif dalam interaksi lisan.

Kerugian

1. Masalahnya yaitu bahwa kemampuan-kemampuan hitrionik dan beberapa kandidat mungkin memperimbnagkan kemurahan mereka atas biaya yang lebih memusatkan perhatian kepada dirinya sendiri. Juga ada masalah pada semua interaksi lisan yaitu bahwa para kandidat sering menggunakan bahasa untuk melaporkan dan mengatakan apa yang akan mereka katakan daripada berperan secara langsung.
2. Paksaan praktek berjalan dengan baik, khususnya dalam ujian-ujian yang bersekala besar. Jika harus menggunakan *role play* yang berbeda maka harus memberikan perhatian besar untuk memastikan bawa mereka berkedudukan sama dengan perminaat-permintaan para kandidat.

BAB VI

TES HASIL BELAJAR BAHASA

6.1 Pengertian Tes Hasil Belajar

Tes sebagai salah satu teknik pengukuran dapat didefinisikan *A test will be defined as a systematic procedure for measuring a sample of an individual's behaviour* (Brown,1970:2). Definisi tersebut mengandung dua hal pokok yang perlu di perhatikan dalam memahami makna tes, yaitu

Pertama adalah kata *systematic procedure* yang artinya bahwa suatu tes harus disusun, dilaksanakan (diadministrasikan) dan diolah berdasarkan aturan-aturan tertentu yang telah ditetapkan. Sistematis di sini meliputi tiga langkah, yaitu (a) sistematis dalam isi, artinya butir-butir soal (item) suatu tes hendaknya disusun dan dipilih berdasarkan kawasan dan ruang lingkup tingkah laku yang akan dan harus diukur atau dites, sehingga tes tersebut benar-benar tingkat validitasnya dapat dipertanggungjawabkan, (b) sistematis dalam pelaksanaan

(administrasi) artinya tes itu hendaknya dilaksanakan dengan mengikuti prosedur dan kondisi yang telah ditentukan ; dan (c) sistematis di dalam pengolahannya, artinya data yang dihasilkan dari suatu tes diolah dan ditafsirkan berdasarkan aturan-aturan dan tolak ukur (norma) tertentu.

Kedua adalah *measuring of an individual's is behaviour* yang artinya bahwa tes itu hanya mengukur suatu sampel dari suatu tingkah laku individu yang dites. Tes tidak dapat mengukur seluruh (populasi) tingkah laku, melainkan terbatas pada isi (butir soal) tes yang bersangkutan.

Suatu tes akan berisikan pertanyaan-pertanyaan dan atau soal-soal yang harus dijawab dan atau dipecahkan oleh individu yang dites (testee), maka disebut tes hasil belajar (*achievement test*). Hal ini sependapat dengan seorang ahli yang menyatakan bahwa *The type of ability test that describes what a person has learned to do is called an achievement test* (Thordike & Hagen, 1975:5). Berdasarkan pendapat itu, tes hasil belajar biasanya terdiri dari sejumlah butir soal yang memiliki tingkat kesukaran tertentu (ada yang mudah, sedang, dan sukar). Tes tersebut harus dapat dikerjakan oleh siswa SMP dalam waktu yang sudah ditentukan. Oleh karena itu, tes hasil belajar merupakan *power test*. Maksudnya adalah mengukur kemampuan siswa SMP dalam menjawab pertanyaan atau permasalahan.

6.2 Jenis dan Bentuk Tes Hasil Belajar

Tes merupakan serangkaian soal yang harus dijawab oleh siswa SMP. Dalam hal ini, tes hasil belajar dapat digolongkan kedalam tiga jenis, yaitu (a) tes lisan, (b) tes tulisan, dan (c) tes tindakan atau perbuatan.

Penggunaan jenis tes tersebut seyogianya disesuaikan dengan kawasan domain tingkah laku siswa SMP yang hendak diukur. Misalnya tes tulisan dan tes lisan dapat digunakan untuk mengukur kawasan kognitif, sedangkan kawasan psikomotor dapat diukur dengan tes perbuatan, dan kawasan apektif biasanya diukur oleh skala penilaian yang biasanya disebut tes skala sikap.

Dalam tes tertulis dapat digunakan beberapa bentuk butir soal, yaitu (1) tes bentuk uraian, yang terdiri dari atas tes uraian terikat dan tes uraian bebas (2) serta tes bentuk objektif, yang terdiri dari data butir soal benar atau salah, pilihan ganda, isian singkat, dan menjodohkan.

a. Soal Bentuk Uraian (Esai)

Bentuk soal ini disebut bentuk uraian, karena peserta tes harus menjawab soal-soalnya dengan uraian yang mempergunakan bahasa sendiri secara lugas. Di samping itu tes uraian merupakan salah satu jenis tes tertulis yang umumnya berupa pertanyaan-pertanyaan yang mengandung permasalahan dan memerlukan pembahasan, uraian, atau penjelasan sebagai jawaban. Ciri tes uraian memberikan kebebasan kepada siswa SMP untuk mengorganisasikan jawabannya. Siswa SMP bebas memilih pendekatan yang dipandang dapat dalam menyelesaikan permasalahan yang ditanyakan serta dalam menyusun jawabannya.

Berdasarkan uraian di atas, 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.

Tes uraian merupakan tes yang tertua, namun bentuk ini masih digunakan secara luas di Amerika Serikat hingga kini, bahkan merupakan bentuk soal yang yang juga masih digunakan secara luas di bagian-bagian dunia lainnya (Gronlund, 1977).

Tes uraian memiliki beberapa kelebihan dibandingkan dengan tes objektif, yaitu (1) memungkinkan para testi menjawab soal secara bebas sepenuhnya, (2) merupakan tes yang terbaik dalam mengukur kemampuan menjelaskan, membandingkan, merangkum, membedakan, menggambarkan, dan mengevaluasi ; (3) merupakan tes yang terbaik untuk mengukur keterampilan mengemukakan pendapat dengan tulisan; (4) memberikan kesempatan bagi siswa SMP untuk meningkatkan kemampuan menulis, mengorganisasikan ide serta berfikir secara kritis dan kreatif ; (5) dapat menggalakan siswa SMP mempelajari secara luas tentang sebagian besar konsep dan menggeneralisasikan; (6) bila dibandingkan dengan bentuk tes yang lain tes uraian relatif lebih mudah membuatnya; (7) secara praktis para siswa SMP tidak mungkin menebak jawaban yang benar; dan (8) mungkin lebih sesuai untuk mengukur kemampuan kognitif yang relatif lebih tinggi (lihat Balitbang Dikbud, 1984 : 24).

Berdasarkan uraian di atas dapat dikatakan bahwa tes uraian dapat dijadikan sebagai suatu alternatif untuk mengatasi dampak yang negatif yang dapat terjadi dalam penggunaan tes objektif. Selain itu, tes uraian mampu mengungkapkan aspek pengetahuan yang kompleks secara mendalam ; mampu melihat jalan pikiran siswa, menuntut siswa SMP untuk mengkreasikan dan mengorganisasikan jalan pikiran mereka dalam jawaban soal.

Tes bentuk uraian memiliki ciri-ciri tertentu, seperti yang dikemukakan oleh Wirasasmita (1981 : 24) yaitu (a) hendaknya setiap pertanyaan merupakan suatu perumusan yang jelas, definitif, dan pasif, (b) tiap pertanyaan hendaknya disertai petunjuk yang jelas tentang jawaban yang dikehendaki oleh oleh peserta, (c) hendaknya pertanyaan-pertanyaan tersebut mencakup semua bahan yang terpenting serta komprehensif, (d) perbandingan soal sukar, sedang, dan mudah harus seimbang, walaupun belum ada patokan yang pasti. Sebaiknya perbandingannya, sukar = 30% - 25%, sedang = 50%, dan mudah = 20% - 25%, dan setelah soal disusun segera susun kunci jawabannya, dengan memperhatikan berbagai kemungkinan jawaban.

b. Tes Bentuk Objektif

Soal bentuk ini bermacam-macam diantaranya adalah

- bentuk benar salah (true false);
- bentuk menjodohkan (matching);
- bentuk isian (completion); dan
- bentuk pilihan ganda (multiple choice) .

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 SMP adalah tes bentuk objektif (1) tepat untuk mengungkapkan hasil belajar yang bertatanan pengetahuan, pemahaman, aplikasi, dan analisis, (2) mempunyai dampak belajar yang mendorong siswa SMP 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) siswa SMP tidak dituntut untuk mengorganisasikan jawaban, karena jawabannya sudah disediakan, (2) siswa SMP ada kemungkinan dapat menebak jawaban yang telah tersedia (3) tidak dapat mengungkap proses berpikir dan bernalar, (4) hanya mengukur ranah kognitif yang paling rendah tidak mengungkap kemampuan yang lebih kompleks. Hal ini sebagaimana yang diungkapkan Gronlund (1985 : 36) menyatakan bahwa *...objective test items can be used to measure a variety of knowledge out come ...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.

Pendapat lain yang berbeda, yakni Lado (1961 : 201) mengemukakan bahwa *The usual objections to objective test are that they are too simple, that they do not require real thinking but simple memory, and that they do not test the ability of the student to organize his thought.*

Pendapat di atas menunjukkan bahwa keberatan tes objektif adalah karena tes itu terlalu mudah, tidak menuntut pemikiran yang nyata, dan tidak menguji kecakapan siswa SMP dalam mengorganisasikan pikirannya. Padahal pada tingkatan perguruan tinggi kemampuan untuk mengorganisasikan pemikiran, mengungkapkan ide secara sistematis, dan menunjukkan kemampuan nalar yang ilmiah merupakan tuntutan yang ditujukan kepada siswa SMP, lebih jauh kepada lulusan perguruan tinggi (Ditjen Dikdasmen, 1982/1983 : 20).

Dilihat dari sudut waktu kapan dan untuk apa tes itu dilakukan, maka tes hasil belajar dapat dikelompokkan menjadi tes awal (*pretest*), tes akhir (*posttest*), dan *entering behaviour test*.

Tes awal biasanya dilakukan setelah proses belajar mengajar selesai. Tujuannya untuk mengetahui tingkat penguasaan mahasiswa SMP terhadap materi pelajaran yang telah diberikan pada proses belajar mengajar yang bersangkutan. Tujuan lain adalah untuk memperbaiki proses belajar mengajar yang telah dilakukan, hasilnya disebut hasil tes formatif, sedangkan bila tujuannya untuk menetapkan lulusan atau kenaikan kelas seseorang terhadap mata pelajaran tertentu maka disebut ujian akhir atau ulangan umum.

Entering behaviour test adalah suatu tes yang berisikan materi pelajaran atau kemampuan-kemampuan siswa SMP yang harus sudah dikuasai sebelum mereka menempuh suatu proses.

6.3 Kompetensi Dasar Berbahasa bagi Siswa

Proses kegiatan belajar mengajar dalam kelas tidak terlepas dari kegiatan penilaian dan pengukuran keberhasilan pembelajaran. Oleh karena itu, dalam Pedoman KBM berdasarkan Kurikulum Berbasis Kompetensi Mata Pelajaran Bahasa Indonesia ini diberikan pula beberapa petunjuk dan pedoman penilaian keberhasilan pembelajaran bahasa. Seperti kita ketahui bahwa perubahan kurikulum merupakan salah satu upaya untuk meningkatkan mutu pembelajara. Perubahan Kurikulum 1994 yang beroreintasi pada pendekatan komunikatif menjadi Kurikulum Berbasis Kompetensi ini pun merupakan suatu upaya penyempurnaan dan perbaikan kualitas pembelajaran. Indikator keberhasilan pembaharuan kurikulum ditandai dengan adanya perbuahan pada pola kegiatan belajar mengajar, memilih media pendidikan, dan menentukan pola penilaian yang menentukan hasil pembelajaran.

Pembaharuan Kurikulum Bahasa Indonesia mulai dari tingkatan pendidikan dasar sampai pendidikan menengah akan bermakna bila diikuti oleh perubahan praktik-praktik pembelajaran di kelas yang dengan sendirinya akan mengubah juga praktik penilaian pembelajaran. Selama ini praktik penilaian di kelas kurang menggunakan metode dan alat yang lebih bervariasi. Oleh karena itu, seorang guru bahasa Indonesia harus mengetahui dan menguasai serta mampu

menyusun tes-tes bahasa untuk mengukur keberhasilan pembelajaran bahasa Indonesia. Di bawah ini diuarikan beberapa petunjuk dan pedoman tentang (a) dimensi-dimensi tes bahasa sebagai instrumen penilaian dan pengukuran, (b) penilaian berbasis kelas, (c) penilaian kompetensi dalam KBK, (d) acuan kriteria dan acuan norma, serta (e) perencanaan dan pengolahan hasil penilaian.

Penilaian adalah suatu proses yang sistematis dalam memperoleh dan mempergunakan informasi untuk membuat pertimbangan yang dipergunakan sebagai dasar pengambilan keputusan. Penilaian bahasa Indonesia yang dilakukan saat ini masih beorientasi pada pengujian teori bahasa dan teori pendidikan bahasa bukan pada aspek penggunaan bahasa.

Richard (1987:49) menjelaskan bahwa kompetensi komunikatif dalam

pembelajaran bahasa meliputi (1) pengetahuan mengenai gramatika dan kosakata,

(2) pengetahuan mengenai kaidah-kaidah berbicara, (3) pengetahuan mengenai bagaimana cara menggunakan dan memberi respons terhadap tindak-tutur, dan (4) pengetahuan mengenai bagaimana cara menggunakan bahasa secara tepat dan memuaskan. Selanjutnya, ia mengemukakan bahwa komponen-komponen kompetensi komunikatif meliputi (1) kompetensi gramatikal, (2) kompetensi sociolinguistik, (3) kompetensi wacana, dan (4) kompetensi strategi.

Bachman (1990:87) menyebutnya bukan kompetensi komunikatif, tetapi kompetensi bahasa (*language competence*), yang meliputi (1) kompetensi organisasi dan kompetensi pragmatik. Kompetensi organisasi diklasifikasi lagi menjadi (a) kompetensi gramatikal dan (b) kompetensi

wacana. Kompetensi pragmatik pun diklasifikasi menjadi (a) kompetensi *illocutionary competence* dan (b) kompetensi sosiolinguistik.

Kompetensi gramatikal menurut Bachman adalah kemampuan berbahasa dalam hal penguasaan dan penggunaan kaidah-kaidah bahasa, seperti kosakata, pembentukan kata, pembentukan kalimat, dan pembentukan bunyi/sistem penulisan. Sedangkan kompetensi wacana (*textual competence*) yaitu kemampuan siswa dalam penggunaan bahasa dalam aspek kekohesifan dan kekoherenan. *Illocutionary competence* mencakupi pemakaian bahasa yang berkaitan dengan fungsi-fungsi bahasa, seperti fungsi regulasi, fungsi heuristik, fungsi ideasional, fungsi imajinasi, fungsi personal, fungsi interpersonal, dan fungsi instrumental. Kompetensi sosiolinguistik meliputi kemahiran berbahasa dalam hal *sensitivity to dialect or variety*, *sensitivity to register*, *sensitivity to naturalness*, dan *references and figures of speech*.

Selanjutnya Littlewood (1981) mengemukakan bahwa ada dua jenis kompetensi komunikatif, yaitu pra komunikatif dan komunikatif. Yang dimaksud dengan kompetensi pra komunikatif adalah kemampuan berbahasa dalam aspek kompetensi struktural dan kuasi komunikatif. Sedangkan kompetensi komunikatif meliputi kemampuan berbahasa dalam aspek komunikatif fungsional dan interaksi sosial.

Untuk dapat menyusun suatu tes bahasa Indonesia yang baik dan terukur, pengembang tes perlu memperhatikan dan mempelajari dimensi-dimensi tes bahasa, yaitu (1) dimensi tujuan tes bahasa itu, (2) dimensi bentuk stimulus tes,

(3) dimensi bentuk respons tes, (4) dimensi isi tes, (5) dimensi kemampuan teres, (6) dimensi teknik tes, dan (7) dimensi reliabilitas dan validitas tes.

Dimensi tujuan tes bahasa meliputi empat jenis, yaitu tes pencapaian atau tes kemajuan, tes sikap, tes diagnostik, dan tes penempatan.

Dimensi bentuk stimulus tes adalah satu runtunan stimulus dan respons. Oleh karena itu, dalam penyusunan tes bahasa perlu memperhatikan bentuk stimulus yang perlu dirancang oleh pengembang tes. Stimulus-stimulus tersebut dapat berupa lisan, tertulis, gambar-gambar, dan tindakan.

Dimensi bentuk respons merupakan jawaban yang diberikan oleh siswa dalam merespons stimulus yang diberikan oleh penguji. Respons yang diberikan oleh siswa dapat berupa lisan, tertulis, gambar-gambar, dan tindakan.

Dimensi isi tes bahasa berupa tes terpecah dan tes terpadu. Tes terpecah merupakan tes yang hendak mengukur kemampuan siswa dalam penguasaan ejaan dan tanda baca atau penguasaan kosa kata tertentu. Tes terpadu mengukur keseluruhan kemampuan siswa berbahasa sesuai dengan jenjang pendidikan dan tujuan pengajaran yang sudah ditetapkan.

Sesuai dengan tujuan pembelajaran bahasa Indonesia yang menekankan aspek komunikatif, maka tes bahasa Indonesia pun harus mengukur kemampuan siswa dalam berbahasa untuk kepentingan komunikasi. Kemampuan komunikatif siswa yang dimaksudkan adalah kemampuan berbahasa Indonesia sesuai dengan situasi dan konteks. Indikator kemampuan yang dites itu adalah tingkat kompetensi komunikatif.

Dimensi teknik tes bahasa meliputi dikte, esei (mengarang), wawancara, pilihan berganda, tes rumpang (*cloze test*), dan terjemahan.

Tes bahasa yang baik adalah tes yang memiliki reliabilitas dan validitas yang tinggi. Sesuai dengan tujuan pembelajaran bahasa Indonesia yang menekankan pada aspek kompetensi dasar berbahasa Indonesia (komunikatif), maka penilaian bahasa Indonesia pun harus mengukur kompetensi dasar berbahasa Indonesia yang sesuai dengan situasi dan konteks pemakaiannya. Secara umum, kompetensi dasar berbahasa Indonesia ini mengintegrasikan antara keterampilan berbahasa dengan aspek kebahasaan dan kesastraan. Untuk lebih jelasnya kemampuan yang harus dievaluasi dapat dilihat pada bagan di bawah ini.

PENILAIAN KEMAMPUAN BERBAHASA INDONESIA

Keterampilan berbahasa	Produktif		Reseptif	
	Menulis (40 %)	Berbicara (10 %)	Membaca (40 %)	Menyimak (10 %)
A. Konsep Kebahasaan				
Fonologi	-	V	V	V
Ejaan	V	-	V	-
Morologi	V	V	V	V
Sintaksis	V	V	V	V
Semantik	V	V	V	V
Wacana	V	V	V	V

Kosa Kata	V	V	V	V
Sastra	V	V	V	V

Kompetensi dasar berbahasa Indonesia yang harus dinilai adalah kompetensi-kompetensi dasar yang ada dalam Kurikulum Berbasis Kompetensi Mata pelajaran Bahasa dan Sastra Daerah (Indonesia), seperti contoh-contoh dalam bab II buku ini. Kompetensi yang dinilai adalah kompetensi berbahasa Indonesia bukan menilai konsep kebahasaan dan kesastraan. Misalnya dalam kita akan menilai kemampuan menulis tentu saja secara tidak langsung menilai konsep konsep ejaan, kosa kata dan semantik, morfologi, sintaksis, serta wacana bahasa Indonesia.

Kemampuan yang dinilai berdasarkan tingkat kompetensi dasar yang harus dimiliki oleh seorang peserta didik. Tingkatan atau level bagi peserta didik yang bersekolah selama 12 tahun dapat dilihat pada bagan di bawah ini.

LEVEL KOMPETENSI DASAR PEMBELAJARAN BAHASA

Level 0	Selesai TK & RA
Level 1	Selesai kelas II SMP & MI (akhir tahun ke-2)
Level 2	Selesai kelas IV SMP & MI (akhir tahun ke-4)
Level 3	Selesai kelas VI SMP & MI (akhir tahun ke-6)
Level 4	Selesai kelas II SMP & MTs (akhir tahun ke-8)
Level 4A	Selesai kelas III SMP & MTs (akhir tahun ke-9)
Level 5	Selesai kelas I SMA & MA (akhir tahun ke-10)
Level 6	Selesai kelas III SMA & MA (akhir tahun ke-12)

Rentang waktu dalam level-level di atas adalah 2 tahun. Rentang waktu ini lebih pendek dari kompetensi tamatan jenjang TK & RA 2 tahun, jenjang SMP & MI 6 tahun, jenjang SMP & MTs 3 tahun, dan jenjang SMA & MA 3 tahun. Rentang waktu yang lebih pendek ini bertujuan untuk memudahkan guru atau sekolah dalam mengetahui tingkat pencapaian siswa pada level tersebut.

Dengan memahami kompetensi siswa lebih dini dalam rentang waktu yang lebih pendek, guru, orang tua, dan staf sekolah lainnya diharapkan dapat memberikan perbaikan-perbaikan sejak dini sebelum terlambat ketika siswa berada pada kelas terakhir untuk mencapai kompetensi tamatan dari suatu jenjang tertentu. Selain itu, penentuan level-level ini pun bermanfaat bagi kepala sekolah dalam menentukan guru-guru strategis pada setiap level.

a. Kemampuan Menyimak

Sesuai dengan namanya, penilaian kemampuan menyimak lebih tepatnya pengujian kompetensi bahasa lisan, bahkan penilaian kemampuan yang diujikan secara lisan dan diterima siswa melalui sarana pendengaran. Kemampuan menyimak dimaksudkan sebagai kemampuan menangkap dan memahami bahasa lisan.

Tujuan dari penilaian menyimak ini meliputi dua macam, yaitu (1) untuk menilai kemampuan membedakan antar fonem dan bukan hanya untuk

memahami pesan verbal saja dan (2) untuk menilai pemahaman menyimak.

Untuk menilai tujuan-tujuan tertentu yang berkaitan dengan kompetensi dasar berbahasa Indonesia secara lisan dapat dilakukan dengan dua cara, yaitu penilaian diskriminasi fonem dan sensitifitas penekanan serta penilaian pemahaman menyimak. Penilaian menyimak dapat dilakukan dengan beberapa, di antaranya adalah dengan penilaian diskriminasi yang terdiri atas sebuah gambar yang disertasi oleh tiga atau empat kata, kemudian diucapkan oleh penguji secara langsung atau melalui tape. Tipe ini biasanya digunakan untuk menilai kemampuan menyimak pada tahapan tingkat rendah.

Secara alami bahasa Indonesia bersipat lisan dan berwujud dalam kegiatan berbicara dan menyimak. Pada kenyataannya berbahasa lisan lebih banyak digunakan oleh penutur bahasa Indonesia. Oleh karena itu, penilaian kemampuan menyimak perlu mendapat perhatian yang memadai walaupun porsiya tidak sama dengan keterampilan berbahasa Indonesia lainnya (membaca dan menulis).

Dalam pelaksanaannya pembelajaran bahasa Indonesia di sekolah, pembelajaran menyimak apalagi penilaiannya kurang mendapat perhatian sebagaimana halnya keterampilan berbahasa Indonesia lainnya. Belum semua guru mengajarkan dan sekaligus menguji kemampuan menyimak muridnya dalam satu periode teretntu.

Masalah yang dihadapi oleh guru dalam pelaksanaan penilaian kemampuan menyimak adalah berupa sarana rekaman atau langsung yang harus dipersiapkan oleh guru dalam penilaian berlangsung. Penggunaan rekaman untuk

pelaksanaan penilaian kompetensi dasar menyimak mempunyai keuntungan , di antaranya yaitu (1) menjamin tingginya tingkat keterpercayaan alat tes, (2) memungkinkan kita untuk membandingkan prestasi antara kelas yang satu dengan kelas yang lain walaupun selang waktu cukup lama, (3) jika alat penilaian memiliki tingkat kesahihan dan keterpercayaan yang memadai, dapat diupergunakan berkali-kali, (4) dapat merekan situasi tertentu pemakaian bahasa Indonesia di masyarakat untuk dibawa ke kelas, serta (5) guru dapat mengontrol pelaksanaan penilaian dengan lebih baik (lihat Nurgiyantoro,1988:231).

Bahan yang perlu diperhatikan dalam menilai kemampuan menyimak adalah (a) tingkat kesulitan wacana, (b) isi dan cakupan wacana, serta (c) jenis-jenis wacana. Tingkat kemampuan menyimak meliputi tingkatan ingatan, pemahaman, penerapan, dan analisis.

Tingkat kesulitan wacana dapat dilihat dari faktor kosa kata dan struktur bahasa yang digunakan. Jika kosa kata yang dipergunakan sulit, bermakna ganda dan abstrak, jarang dipergunakan, dan ditambah lagi struktur kalimatnya juga kompleks, wacana tersebut termasuk wacana yang tinggi tingkat kesulitannya. Akan tetapi, jika kedua aspek kebahasaan tersebut sederhana, wacana tersebut tergolong wacana sederhana. Ada suatu cara untuk memperkirakan tingkat kesulitan wacana bagi kelas, yaitu berupa *cloze (cloze test)*. Teknik ini diberikan secara lisan (*oral cloze procedure*). Caranya wacana dibaca oleh guru (penguji) di depan kelas dua klai, dan setiap pada kata yang ke-n (ke-5, ke-6 atau ke-7) tidak dibaca. Siswa diminta untuk menerka dan kemudian menuliskan kata-kata yang

tidak dibaca tersebut pada secarik kertas. Jika rata-rata jawaban betul siswa kurang atau hanya mencapai 20%, wacana yang bersangkutan termasuk wacana yang sulit bagi siswa di kelas tersebut. Sebaliknya, jika jawaban betul siswa minimal 75%, wacana tersebut tergolong mudah bagi kelas yang bersangkutan. Wacana yang baik untuk dipergunakan dalam penilaian kemampuan menyimak adalah wacana yang tidak terlalu sulit atau sebaliknya terlalu mudah (band. Nurgiyantoro, 1988:233).

Isi dan cakupan wacana biasanya mempengaruhi tingkat kesulitan wacana. Jika isi atau cakupan wacana itu sesuai dengan minat dan kebutuhan siswa atau sesuai dengan bidang yang dipelajari, hal itu akan mempermudah wacana yang bersangkutan. Sebaliknya, jika isi wacana itu tidak sesuai dengan minat dan kebutuhan siswa, ia akan menambah tingkat kesulitan wacana yang bersangkutan. Wacana bahasa Indonesia yang akan dinilai hendaknya berisi hal-hal yang bersikap netral sehingga dimungkinkan adanya kesamaan pandangan terhadap isi wacana itu. Jenis wacana yang dijadikan bahan penilaian menyimak berupa sebuah dialog atau monolog (narasi, deskripsi, argumentasi, eksposisi, ceramah, dan lain-lain).

Tingkat kemampuan menyimak jenjang ingatan hanya sekedar menuntut siswa untuk mengingat fakta atau menyebutkan kembali fakta-fakta yang terdapat di dalam wacana yang telah diperdengarkan sebelumnya. Fakta itu berupa nama, peristiwa, angka, tanggal, tahun, dan sebagainya. Bentuk soal yang digunakan dapat berupa bentuk objektif isian singkat atau pilihan ganda.

Tingkat kemampuan menyimak jenjang pemahaman menuntut siswa untuk dapat memahami wacana yang diperdengarkan. Kemampuan pemahaman ini dimaksudkan siswa harus memiliki pengetahuan tentang isi wacana, hubungan antaride, antarfaktor, antarkejadian, hubungan sebab-akibat, dan sebagainya.

Tingkat kemampuan menyimak jenjang penerapan dimaksudkan agar siswa memiliki kemampuan menerapkan konsep atau masalah tertentu pada situasi yang baru. Butir-butir kemampuan menyimak yang dapat dikategorikan penilaian tingkat penerapan adalah butir soal yang terdiri dari pernyataan (diperdengarkan) dan gambar-gambar sebagai alternatif jawaban yang terdapat di dalam lembar tugas. Siswa menyimak sebuah wacana (kalimat) satu kali dan tugas siswa adalah memilih di antara beberapa gambar yang disediakan yang sesuai dengan wacana.

b. Kemampuan Membaca

Sejumlah definisi *membaca* telah disampaikan oleh para pakar pengajaran membaca. Namun seperti juga keterampilan berbahasa lain, pada dasarnya membaca adalah proses komunikasi, terutama antara teks tertulis (gagasan penulis) dan pembaca. Dalam hal ini keberhasilan membaca akan sangat bergantung pada keberhasilan komunikasi itu sendiri. Salah satu definisi *membaca* yang dapat diterima secara luas adalah “Membaca adalah sebuah proses interaktif antara pengetahuan awal pembaca tentang isi bacaan dan tujuan membaca sehingga mempengaruhi apa yang dipelajari dari teks” (McKenna & Robinson, 1993: 21).

Dalam menjelaskan proses membaca ini, selanjutnya McKenna & Robinson (1993) menyatakan bahwa *reading is defined as the reconstruction in the mind of meaning encoded in print* (membaca dapat pula dikatakan sebagai rekonstruksi makna di dalam pikiran pembaca)

Faktor-faktor afektif, kognitif, dan linguistik saling berinteraksi dalam membentuk dan mempengaruhi kemampuan membaca seseorang. Dalam sebuah penelitian Athey (1985) mengungkapkan beberapa faktor afektif yang mempengaruhi kemampuan membaca: konsep diri, kemandirian, penguasaan lingkungan, persepsi tentang realitas dan kecemasan.

Dalam konteks kognisi aspek-aspek memori sangat penting dalam perkembangan kemampuan membaca. Memori ini terdiri atas memori jangka pendek dan memori jangka panjang. Namun apa yang sangat penting bagi kognisi adalah kemampuan individu dalam membentuk konsep. Menurut Alexander (1988: 8), konsep adalah sekumpulan stimulus yang memiliki karakteristik yang sama. Pembentukan konsep ini sangat penting untuk berpikir dan membaca.

Faktor penting lain yang berkaitan dengan fungsi kognitif adalah metakognisi. Metakognisi ini adalah istilah yang digunakan untuk menjelaskan pengetahuan seseorang tentang ciri-ciri proses berpikirnya dan pengaturan pemikirannya. Jika seseorang memiliki kesadaran metakognitif, maka membaca akan menjadi proses berpikir yang aktif dan

pemahaman pun akan mudah dicapai. Istilah lain yang digunakan untuk menjelaskan fungsi kognitif ini adalah skemata. Menurut Rumelhart (1980), skemata adalah fungsi di dalam otak yang menafsirkan, mengatur dan menarik kembali informasi; dengan kata lain, skemata adalah kerangka mental. Skemata ini sangat penting untuk proses belajar membaca karena skemata menyimpan data masa lalu (pengetahuan dan pengalaman) di dalam memori, yang sewaktu-waktu dapat ditarik kembali jika diperlukan.

Faktor ketiga yang juga sangat penting adalah kemampuan berbahasa. Karena membaca bergantung pada bahasa, maka kemampuan berbahasa seseorang akan mempengaruhi kemampuan membacanya. Namun, membaca berbeda dengan menyimak atau berbicara (DeStefano, 1981). Membaca lebih menuntut kemampuan berbahasa si pembaca karena ia harus bergantung pada bahan bacaan saja atau pada kata-kata tertulis saja, sedangkan bahasa tertulis seringkali lebih kompleks daripada bahasa lisan. Di samping, membaca menuntut seorang pembaca untuk menguasai kaidah-kaidah itu, fonologis, morfologis, sintaksis, dan semantik.

Dari apa yang dikemukakan di atas, dapat dikatakan bahwa membaca adalah sebuah proses kompleks yang membuat pengajarannya sebagai proses yang kompleks pula. Namun, guru membaca yang baik mempunyai satu hal yang sama, yaitu mereka berpikir tentang membaca. Hal ini tidak berarti bahwa semua guru membaca yang baik mempunyai pikiran yang sama. Banyak guru membaca yang baik tidak memiliki pengetahuan atau

preferensi tertentu tentang teori proses membaca atau teori pengajaran membaca. Apa yang membedakan mereka adalah kecenderungan untuk memikirkan peranan mereka dalam pengajaran membaca, untuk mengembangkan pendekatan personal terhadap pengajaran membaca yang menggabungkan apa yang mereka ketahui tentang proses membaca, tentang diri mereka sebagai guru, tentang pengajaran membaca dan tentang pembelajar yang mereka ajari.

Sementara itu, Otto et. al (1979: 4) mengakui bahwa proses membaca dan pengajaran membaca memang begitu kompleks, sehingga para ahli dapat memantaunya dari berbagai sudut pandang. Sedikitnya ada lima disiplin ilmu yang dapat memberikan penjelasan tentang bagaimana proses membaca berlangsung. Disiplin ilmu pertama adalah psikologi, yang mengkaji proses ini melalui pendekatan perseptual/konseptual, behavioristik, nativistik, kognitif, dan psikometrik. Psikolinguistik adalah disiplin ilmu kedua yang juga memberikan kontribusi terhadap pemaparan proses membaca. Bidang pengolahan informasi (*information processing*) adalah bidang ketiga yang mengkaji proses membaca dari sudut pandang sibernetika, analisis sistem dan teori komunikasi umum. Sosiolinguistik adalah bidang ilmu keempat yang memberikan kontribusi terhadap pemahaman tentang proses membaca dan khususnya tentang proses pengajaran membaca. Terakhir, ilmu-ilmu perilaku juga membantu meningkatkan wawasan dan pemahaman tentang aspek-aspek tertentu dalam proses membaca.

Selain itu, para teoritikus mendekati proses membaca dengan berbagai cara dan sudut pandang yang berbeda. Misalnya, ada beberapa jenis teori: teori makro dan teori mikro. Teori makro berusaha membahas kegiatan membaca dalam seluruh kompleksitasnya. Sedangkan teori mikro dirancang untuk menjelaskan satu segmen kecil dalam proses membaca. Selain itu, ada pula teori perkembangan dan teori deskriptif. Teori perkembangan adalah upaya untuk menjelaskan kegiatan membaca menurut cara proses membaca itu dipelajari, sedangkan teori deskriptif berusaha mendeskripsikan tindakan-tindakan pembaca yang proses membaca. Terakhir, ada pendekatan molekuler dan pendekatan holistik terhadap pengembangan kemampuan membaca. Pendekatan molekuler berusaha menguraikan proses membaca ke dalam perilaku-perilaku atau keterampilan-keterampilan tertentu dan menunjukkan bagaimana semua perilaku ini digabungkan dalam mencapai keberhasilan membaca. Sebaliknya, pendekatan holistik kurang menekankan perilaku-perilaku tertentu, tetapi lebih menitikberatkan pada hubungan atau keterkaitan yang kompleks di antara komponen-komponen proses membaca.

Salah satu skala kemampuan membaca pemahaman itu disusun oleh Departemen Pendidikan Amerika Serikat yang membidangi National Assessment of Educational Progress adalah sebagai berikut.

1) Rudimentary (Sangat Dasar)

Pembaca yang telah menguasai keterampilan dan strategi membaca yang sangat dasar dapat mengikuti petunjuk tertulis yang singkat. Mereka juga dapat memilih kata, frase, atau kalimat untuk menjelaskan sebuah gambar sederhana dan dapat menafsirkan isyarat-isyarat sederhana untuk mengenal sebuah objek biasa. Kemampuan pada taraf ini menunjukkan kemampuan untuk melakukan tugas-tugas membaca yang sederhana.

2) *Basic* (Dasar)

Pembaca yang telah mempelajari keterampilan dan strategi pemahaman dasar dapat menemukan, dan mengenali fakta-fakta dari paragraf informasi, cerita dan artikel berita sederhana. Di samping itu, mereka dapat menggabungkan berbagai gagasan dan menarik kesimpulan yang didasarkan pada bahan bacaan pendek. Kemampuan pada taraf ini adalah memahami informasi spesifik.

3) *Intermediate* (Menengah)

Pembaca yang memiliki kemampuan dan strategi ini dapat mencari, menemukan, dan menyusun informasi yang ada dalam bahan bacaan yang relatif panjang dan dapat membuat parafrase dari apa yang telah mereka baca. Mereka juga dapat menarik kesimpulan dan mencapai generalisasi tentang gagasan utama dan tujuan penulis. Pada taraf ini kemampuan yang dikuasai adalah mencari informasi spesifik, menghubungkan berbagai gagasan, dan membuat generalisasi.

4) *Adept* (Terampil)

Pada taraf terampil ini pembaca dapat memahami teks sastra dan informasi yang rumit, termasuk bahan bacaan tentang topik-topik yang mereka pelajari di sekolah. Mereka juga dapat menganalisis bahan bacaan serta memberikan reaksi atau penjelasan tentang teks secara menyeluruh. Dengan kata lain, mereka dapat menemukan, memahami, merangkum, dan menjelaskan informasi yang relatif kompleks.

5) *Advanced* (Mahir)

Pembaca yang menggunakan keterampilan dan strategi membaca mahir ini dapat mengembangkan dan membentuk kembali gagasan-gagasan yang disajikan dalam teks yang kompleks. Mereka juga mampu memahami hubungan di antara gagasan-gagasan sekalipun hubungan itu tidak dinyatakan secara eksplisit, dan bahkan membuat generalisasi yang tepat meskipun teks tidak memuat keterangan yang jelas.

Menurut Broughton (dalam Tarigan, 1987:11-12) tingkat kemampuan membaca pemahaman terdiri atas dua jenis, yaitu kemampuan yang bersifat mekanik dan kemampuan membaca yang bersifat pemahaman. Kemampuan membaca yang bersifat mekanik merupakan keterampilan membaca tingkat rendah. Indikator atau penanda yang dapat digunakan untuk menentukan apakah seseorang pembaca berada pada tingkat mekanik ini adalah sebagai berikut.

- 1) Pengetahuan pembaca baru sekedar mengenal bentuk-bentuk huruf, angka, dan tanda-tanda yang lain.

- 2) Pembaca baru mengenal bentuk-bentuk linguistik, misalnya: fonem/grafem, kata, frase, klausa, dan kalimat.
- 3) Pembaca baru mengenal hubungan/korespondensi pola ejaan dan bunyi, atau hanya sekedar mampu menyuarakan apa yang ditulis.
- 4) Biasanya kecepatan membaca masih lambat.

Keterampilan membaca pemahaman merupakan kelanjutan dari keterampilan membaca mekanik. Pada tingkat ini, kepada pembaca tidak hanya dituntut untuk mampu mengenal dan membaca unsur-unsur linguistik, melainkan lebih dari itu.

Penelitian ini dilakukan di perguruan tinggi, yaitu dengan tingkat pemahaman bacaan. Aspek-aspek yang hendak dicapai pada taraf membaca tingkat pemahaman adalah seperti berikut.

- 1) Pembaca memahami pengertian-pengertian sederhana dalam hal leksikal (kata-kata), gramatikal (kalimat), dan retorikal (wacana).
- 2) Pembaca dapat memahami signifikansi dan makna yang dibaca.
- 3) Pembaca mampu mengevaluasi bacaan, misalnya evaluasi dari segi bentuk, isi, tanda baca, dan lain-lain.
- 4) Pembaca mampu mengukur kecepatan membacanya, dalam arti pembaca mengetahui kapan ia harus membaca hati-hati, kapan ia harus membaca cepat atau membaca sekilas.

Membaca tingkat pemahaman sangat diperlukan di dalam dunia pendidikan, terutama untuk jenjang perguruan tinggi. Menurut Herbert H. Clark dan Eva V. Clark (1977:43) membaca pemahaman merupakan suatu

proses pembentukan interpretasi atau pengertian. Pemahaman lahir setelah pembaca mengerti apa yang dibacanya. Pengertian ini merupakan jawaban atas pertanyaan yang diajukan pada bacaan. Sejalan dengan pendapat di atas, Smith (1982:62) mengemukakan bahwa pemahaman berarti jawaban-jawaban yang diperoleh dari pertanyaan-pertanyaan yang diajukan terhadap suatu bacaan. Dari kedua pendapat tersebut, dapat disimpulkan bahwa pemahaman merupakan suatu kegiatan membaca untuk memperoleh pengertian yang mendalam dari informasi yang disampaikan penulis. Pengertian yang baik ini akan memudahkan pembaca untuk menginterpretasikan dan menilai permasalahan yang terdapat dalam bacaan, sehingga apabila diajukan pertanyaan-pertanyaan kepada pembaca tersebut dengan mudah akan mudah dijawabnya. Lebih dari itu, pemahaman terhadap suatu bacaan dapat menimbulkan perubahan-perubahan yang positif dari pembaca, baik perubahan dalam bentuk pengetahuan, sikap maupun perubahan dalam bentuk keterampilan.

Menurut Barret (dalam Dupuis, 1982:25-27), pada dasarnya tingkat pemahaman seseorang terhadap bacaan dapat diklasifikasikan atas beberapa tingkat:

- 1) kemampuan mengingat atau memahami kata-kata secara harfiah;
- 2) kemampuan membentuk pengertian (apresiasi) berdasarkan pemahaman di atas;
- 3) kemampuan menarik kesimpulan; dan
- 4) kemampuan mengadakan evaluasi.

Berdasarkan pendapat Barret tersebut, terlihat bahwa kegiatan membaca pemahaman sangat perlu dilakukan untuk mengungkapkan makna dari seluruh bacaan. Melalui kegiatan membaca pemahaman maka dengan mudah kita dapat memperoleh gagasan dan pesan yang terdapat dalam bacaan, sehingga dengan mudah pula pembaca mampu menghubungkan-hubungkan gagasan yang satu dengan gagasan yang lain.

Sejalan dengan pendapat Barret, Gray (dalam Gardner, 1978:65-81) mengemukakan beberapa tingkatan pemahaman terhadap bacaan. Tingkat pemahaman bacaan tersebut dapat diklasifikasikan atas lima tingkatan, yaitu berikut di bawah ini.

- 1) Persepsi awal yang terdiri dari (a) pemahaman terhadap kosakata, (b) pengenalan struktur bacaan, (c) memahami dan mengikuti petunjuk yang terdapat dalam bacaan.
- 2) Pemahaman atau interpretasi terhadap bacaan yang terdiri dari (a) merasakan atau mengetahui tujuan yang hendak dicapai penulis, (b) menemukan hubungan sebab akibat yang terdapat dalam bacaan, (c) mengetahui suasana dan perasaan penulis, (d) menganalisis karakter dan motif yang terdapat dalam bacaan, (e) mencatat kriteria-kriteria dan hubungan-hubungan yang terdapat dalam bacaan, (f) membuat kesimpulan bacaan, dan (g) mampu dan mau berspekulasi dengan peristiwa dan kenyataan.
- 3) Mengadakan evaluasi, yaitu mengukur seberapa jauh pembaca dapat menilai baik tidaknya bacaan yang dibacanya.
- 4) Memberikan reaksi terhadap apa yang dibacanya. Reaksi ini dapat bersifat

emosional intelektual (penuh pertimbangan baik buruk).

5) Mengadakan integrasi bacaan dengan latar belakang pembaca.

Berhasil tidaknya seseorang dalam melakukan kegiatan membaca pemahaman dapat dilakukan dari berbagai hal, yaitu berdasarkan kemampuan mengungkap kembali apa yang telah dibacanya, kemampuan memberikan penilaian terhadap permasalahan yang dikemukakan penulis, kemampuan menerapkan petunjuk-petunjuk yang terdapat dalam bacaan, kemampuan menjawab pertanyaan-pertanyaan yang berkaitan dengan bacaan. Bila pembaca mampu menjawab pertanyaan yang diajukan kepadanya dengan baik, besar kemungkinan baik pulalah pemahaman pembaca tersebut. Demikian pula sebaliknya, banyak aspek yang dinyatakan untuk mengetahui tingkat pemahaman seseorang dalam membaca.

Farr (1969:3) menyatakan bahwa aspek-aspek membaca pemahaman meliputi (1) faktor verbal umum atau *a general verbal factor*, (2) pemahaman eksplisit bahan yang dinyatakan atau *comprehension of explicitly stated material* (3) pemahaman implisit arti yang sebenarnya atau *comprehension of implicit of latent meaning*, dan (4) apresiasi atau *appreciation*. Sementara Tierney (1990:235) menjelaskan bahwa aspek membaca pemahaman meliputi (1) tingkat literal atau *literal level*, (2) tingkat interpretasi atau *interpretative level*, dan (3) tingkat penerapan atau *applied level*.

Selain itu, Dubois (1972:24) mengemukakan taksonomi B Barret membagi tingkatan membaca pemahaman menjadi tingkat lateral, tingkat inferensial, tingkat evaluasi, dan tingkat apresiasi. Pemahaman lateral membutuhkan ingatan pada gagasan-gagasan, informasi, kegiatan-kegiatan yang dinyatakan secara jelas pada bahan bacaan. Pemahaman inferensial merupakan pemahaman yang ditunjukkan ketika pembaca menggunakan sintesis pada isi lateral tersebut pada suatu seleksi, pengetahuan personalnya, intuisinya, dan imajinasinya sebagai suatu dasar untuk penghubung-penghubung hipotesis. Pemahaman evaluasi merupakan pemahaman yang ditunjukkan ketika pembaca menilai suatu bacaan. Pemahaman apresiasi adalah pemahaman yang bersangkutan dengan kesadaran akan teknik-teknik sastra, bentuk, gaya, dan struktur yang digunakan penulis untuk membangkitkan respon-respon emosional pembacanya.

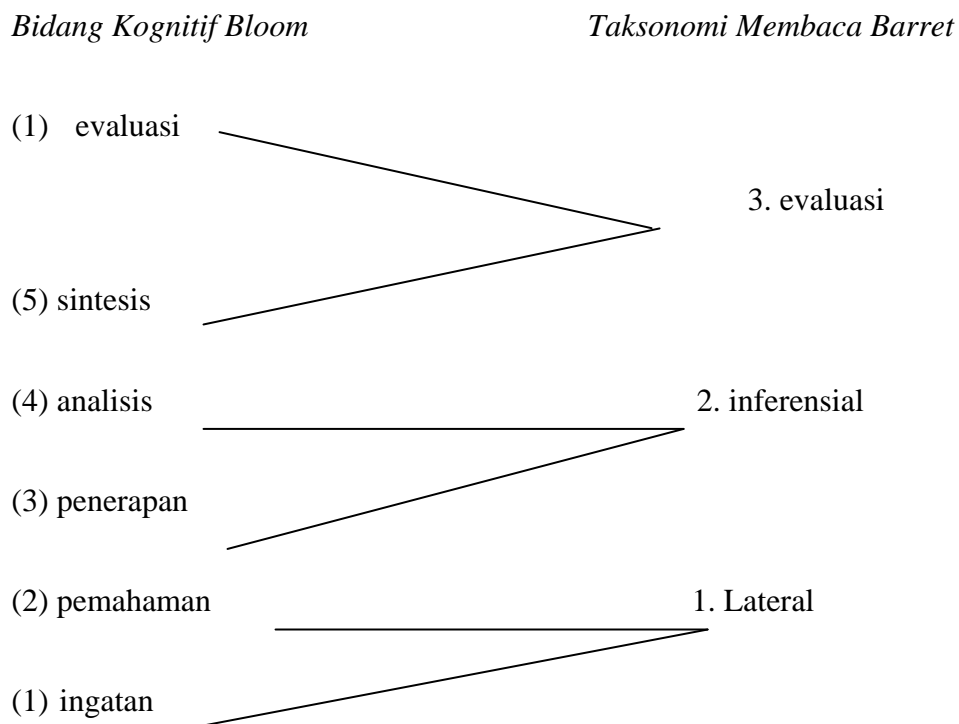
Sebenarnya, Sheila (1982:41) telah mengemukakan aspek-aspek membaca pemahaman yang lebih lengkap, yaitu (1) lateral; (2) reorganisasi; (3) apresiasi; (4) evaluasi; (5) ekstrapolasi. Pemahaman lateral adalah pengidentifikasian dan pengingatan rincian-rincian, ide-ide, fakta, pendapat, konsep, instruksi, contoh, simpulan, dan petunjuk-petunjuk. Pemahaman reorganisasi berisi identifikasi ide-ide dan rekonstruksinya ke dalam ringkasan dan abstrak. Pemahaman apresiasi berupa penghayatan terhadap gaya, perasaan, nuansa-nuansa. Pemahaman evaluasi merupakan tafsiran pendapat, argumen, kritik, dan uraian isi yang

disampaikan. Pemahaman ekstrapolasi adalah kesimpulan di luar wacana, penerapan pada situasi lain.

Penadapat tersebut di atas senada dengan tingkatan membaca pemahaman yang diajukan oleh Barret yang terkenal dengan sebutan Taksonomi Barret". Dupuis (1972:24-28) menyebutkan keempat tingkatan membaca pemahaman itu, yakni (1) pemahaman literal, (2) pemahaman inferensial, (3) pemahaman evaluasi, dan (4) pemahaman apresiasi.

Selanjutnya, Dupuis mengemukakan bahwa pemahaman lateral adalah pemahaman yang membutuhkan ingatan mengenai gagasan-gagasan, informasi, kejadian-kejadian yang dinyatakan secara jelas pada bahan bacaan. Pemahaman inferensial merupakan pemahaman yang ditujukan ketika pembaca menggunakan sintesis pada isi lateral tersebut pada suatu seleksi, pengetahuannya personalnya, intuisi, dan imajinasinya sebagai suatu dasar untuk penghubung-penghubung hipotesis. Pada pemahaman inferensial ini, pernyataan-pernyataan imajinasi memerlukan pemikiran. Evaluasi merupakan yang ditunjukkan ketika pembaca menilai isi bacaan. Ia membandingkan kriteria eksternal dan internal. Kriteria eksternal ditunjukkan dari subjektivitas pengarang dan internal berdasarkan pengalaman pembaca, pengetahuannya yang menghubungkan antara yang ditulis dengan pembaca. Apresiasi adalah pemahaman yang berkaitan dengan kesadaran teknik sastra, bentuk, gaya, dan struktur yang dikerjakan pengarang untuk mendorong respon-respon emosional pembacanya.

Ketiga tingkatan Barret yaitu literal, inferensial, dan evaluasi berhubungan dengan taksonomi Bloom. Pada tingkatan keempat (apresiasi), taksonomi Barret berhubungan dengan tingkat afektif Bloom karena respon dari pembaca terhadap apa yang terkandung dalam bacaan. Jika dibandingkan antara ranah kognitif Bloom dan taksonomi Barret dapat dilihat sebagai berikut.



Gray dalam Keith Gardner (1978:65-81) mengklasifikasikan tingkat pemahaman dalam membaca menjadi lima, tingkat seperti berikut.

- (1) Persepsi awal, yang meliputi: (a) memahami kata, (b) mengenal struktur, (c) membuat ringkasan, dan (d) mengikuti dan memahami petunjuk yang ada dalam bacaan.
- (2) Pemahaman atau interpretasi yang mencakup: (a) merasakan atau mengetahui tujuan pengarang, (b) menemukan hubungan kausal, (c) mengetahui suasana hati perasaan pengarang, (d) menganalisis karakter dan motif, (e) mencatat kriteria dan hubungan, (f) membuat kesimpulan, dan (g) berspekulasi antara peristiwa dan kenyataan.
- (3) Evaluasi, yakni bagaimana pembaca dapat menilai baik tidaknya teks bacaan.
- (4) Reaksi, baik emosional maupun intelektual dari pembaca.
- (5) Integrasi, antara bacaan dan latar belakang pembaca.

Pada pendapat Gray di atas terlihat bahwa pemahaman tidaklah terbatas hanya mengerti makna harfiah yang disampaikan, tetapi hendaknya pembaca tahu tujuan pengarang dan dapat merasakan suasana hati dan perasaan yang dikandung dalam karangan itu. Pembaca dapat bereaksi, baik secara emosional maupun intelektual terhadap materi yang dibacanya.

Smith (1973:231-234) mengemukakan bahwa aktivitas pemahaman membaca dapat dikategorikan menjadi tiga tingkatan, yaitu (1) literal, (2) inferensial, dan (3) evaluasi. Tahap literal pembelajar diharapkan dapat memparafrasekan arti arti yang diberikan dengan jelas dalam wacana. Tahap inferensial merupakan tahap pencarian beberapa jenis organisasi

dari bahan dan mencari ide-ide pada informasi yang ada dalam bacaan. Pada tahap elaboratif, pemahaman yang diharapkan dari pembelajar adalah proses berfikir baru. Penekanannya di sini pada daya kreatif yang dimiliki pembelajar.

Berdasarkan uraian di atas, tingkatan-tingkatan pemahaman secara garis besar ada tiga, yaitu (1) literal, (2) inferensial, dan (3) ekstrapolasi. Peningkatan literal merupakan tingkat pemahaman yang menanyakan apa yang dimaksud oleh pengarang atau pembaca dapat menjelaskan makna secara jelas pada teks atau yang tersurat. Tingkatan inferensial adalah tingkatan pemahaman yang menanyakan apa yang dimaksud oleh pengarang atau pembaca dapat menerapkan organisasi dari bahan dan mencari ide-ide pada bacaan itu. Ekstrapolasi merupakan tingkatan pemahaman yang menyatakan di luar bacaan. Pada tingkatan ini pembaca membuat kesimpulan di luar wacana, kreasi ide-ide dan konsep-konsep serta kesimpulan-kesimpulan lebih jauh dari bacaan.

Kemampuan membaca adalah kemampuan mental pembaca dalam hal memahami apa yang dituturkan pihak lain melalui sarana tulisan. Dalam kemampuan membaca diperlukan pengetahuan tentang sistem penulisan.

Penilaian kemampuan membaca dikaksudkan untuk mengukur tingkat kompetensi dasar siswa dalam memahami wacana tertulis. Kemampuan membaca dapat diartikan sebagai kemampuan untuk memahami informasi yang disampaikan pihak lain melalui sarana tulisan. Kemampuan membaca dapat

diklasifikasikan ke dalam beberapa jenis, yaitu kemampuan membaca pemahaman, membaca cepat, dan membaca indah/teknis. Dalam menyusun alat penilaian membaca, sebaiknya guru dapat melakukan memperhatikan (a) bahan tes kemampuan membaca, yang meliputi tingkat kesulitan wacana, isi wacana, panjang-pendek wacana, dan bentuk wacana serta (b) tingkatan kemampuan membaca. Aspek yang dinilai dalam membaca indah/teknis, di antaranya adalah ketepatan melafalkan bunyi bahasa, ketepatan menggunakan intonasi, keindahan bunyi, dan sebagainya. Aspek yang dinilai dalam membaca cepat adalah jumlah kosa kata, lama waktu membaca dan tingkat kemampuan membaca pemahaman, dengan menggunakan rumus sebagai berikut.

$$KEM = \frac{\text{Jumlah Kosa Kata}}{\text{Waktu membaca}} \times \frac{\text{Skor Tercapai}}{\text{Skor Ideal}} = \dots\dots\dots \text{kata/menit}$$

Dengan mengidentifikasi beberapa kemampuan membaca secara spesifik, ada beberapa tingkatan kemampuan membaca yaitu (a) mengenal kata dan kelompok kata, mengasosiasikan bunyi dengan keterkaitannya pada simbol; (b) menyimpulkan makna suatu kata dengan memahami bentuk kata (akar kata, imbuhan (*rarangken*), derivasi, dan gabungan kata) dan dengan memperhatikan konteks pemakaian bahasa; (c) memahami informasi yang tersirat; (d) memahami hubungan yang berada dalam kalimat, terutama unsur dari struktur kalimat, kata ingkaran, pembukaan

dan tema, sisipan kompleks; (e) memahami hubungan antara bagian-bagian sebuah teks secara mendalam baik dalam hal leksikal (misalnya: dalam *rajekan, kecap saharti, jeung kecap sabalikna*) maupun keterpaduan dalam hal gramatikal terutama referensi anaproik dan kataporik (misalnya, *manehna, maranehna, itu, sok sanajan*); (f) memahami makna konseptual, terutama jumlah dan kuantitas kepastian dan ketidakpastian, perbandingan dan tingkatan, arti dan alat, sebab, hail, maksud, alasan, kondisi, penambah, dan penjelas; (g) mengantisipasi dan memprediksi apa yang akan muncul kemudian dalam teks selanjutnya; (h) mengidentifikasi pikiran utama dan pikiran penjelas; (I) memahami informasi yang tersurat; (j) menggambarkan secara umum dan menarik kesimpulan; (k) menyaring dan mendeteksi (mencari makna secara keseluruhan dan membaca informasi yang spesifik; (l) membaca kritis.

Dalam pemilihan bahan penilaian kemampuan membaca meliputi (a) tingkat kesulitan wacana, (b) isi wacana, (c) panjang-pendek wacana, (d) bentuk-bentuk wacana, dan (e) tingkat-tingkat kemampuan membaca bahasa Indonesia.

Tingkat kesulitan wacana terutama ditentukan oleh kekompleksan kosa kata dan struktur. Semakin sulit dan kompleks kedua aspek itu akan semakin sulit wacana tersebut. Secara umum kita menganggap bahwa wacana yang baik untuk bahan penilaian kemampuan membaca adalah wacana yang tidak terlalu sulit dan tidak terlalu mudah dan yang lebih penting sesuai dengan tingkat kemampuan murid.

Tingkat kesulitan kosa kata umumnya dipergunakan untuk menentukan tingkat kesulitan wacana. Kesulitan kosa kata itu sendiri ditentukan berdasarkan frekuensi pemakaian kosa kata itu dalam wacana. Selain itu, tingkat kesulitan kosa kata pun ditentukan oleh jumlah kosa kata yang digunakan dalam wacana tersebut.

Prosedur pengujian tingkat kesulitan wacana yang dapat dilakukan oleh guru sendiri adalah dengan teknik *cloze*. Wacana yang akan diuji tingkat kesulitannya ditekankan dalam bentuk *cloze test*. Jika rata-rata jawaban betul lebih dari 75%, wacana yang bersangkutan dinyatakan mudah. Sebaliknya, jika rata-rata betul kurang dari 20%, wacana tersebut tergolong sulit bagi siswa yang bersangkutan.

Isi wacana yang dijadikan bahan penilaian kemampuan membaca secara paedagogis harus sesuai dengan tingkat perkembangan jiwa, minat, kebutuhan atau menarik perhatian murid. Kesesuaian tersebut dibutuhkan karena tujuan dari membaca itu sendiri adalah untuk memperluas dunia murid, memperkenalkan berbagai hal dan budaya dari berbagai pelosok daerah. Selain itu, melalui pembelajaran membaca sebenarnya kiat dapat berperan serta mengembangkan sikap dan nilai-nilai pada diri murid, misalnya menyediakan wacana yang berkaitan dengan tata krama, adat istiadat, sejarah perjuangan bangsa, dan sebagainya. Dengan demikian, pemilihan isi wacana perlu disesuaikan dengan perkembangan, minat, sikap, motivasi, dan kebutuhan anak dalam kehidupan di masyarakat.

Panjang–pendek wacana merupakan hal yang penting dalam pemilihan bahan penilaian kemampuan membaca. Wacana yang ditekankan sebaiknya tidak terlalu panjang. Beberapa wacana yang pendek lebih baik daripada sebuah wacana yang panjang, sepuluh butir dari tiga atau empat wacana lebih baik daripada hanya dari sebuah wacana panjang. Keuntungan dengan wacana pendek ini adalah kita dapat membuat soal tentang berbagai hal, lebih komprehensif, serta secara psikologis murid pun lebih senang pada wacana yang pendek, karena tidak membutuhkan waktu yang banyak untuk membacanya dan wacana pendek itu lebih mudah.

Yang dimaksud dengan wacana pendek adalah wacana yang terdiri satu atau dua alinea atau kira-kira sebanyak 50 sampai 100 kata. Wacana pendek bahkan dapat berupa satu kalimat, atau satu pernyataan, yang kemudian dibuat parafrasanya. Penilaian kemampuan membaca dalam hal ini adalah memahami dan memilih parafrase tersebut yang sesuai dengan pernyataan.

Bentuk wacana yang dipergunakan sebagai bahan untuk penilaian kemampuan membaca dapat berbentuk prosa, puisi, dan drama. Umumnya wacana yang dipergunakan berbentuk prosa.

Tingkat kemampuan membaca ditekankan pada kemampuan untuk memahami informasi yang terkandung dalam wacana. Kegiatan memahami informasi itu sendiri sebagai suatu aktivitas kognitif, yaitu

tingkatan pemahaman bacaan dalam jenjang ingatan, pemahaman, aplikasi, analisis, sintesis, dan evaluasi.

c. Kemampuan Menulis

Ada dua metode yang sering digunakan dalam pengukuran kemampuan menulis atau mengarang, yaitu metode langsung dan metode tidak langsung. Metode langsung merupakan tes keterampilan menulis langsung dilaksanakan dengan cara pelaksana tes (guru) langsung menyuruh siswa atau peserta tes menulis atau mengarang topik-topik atau judul-judul karangan tertentu. Keunggulan metode langsung adalah (1) dapat mengukur kemampuan tertentu (kemampuan menyusun, menghubungkan serta memakai bahasa yang dikarangnya dapat lebih efektif, (2) mempunyai potensi untuk mendorong peserta mengerjakan tugasnya sebaik-baiknya; dan (3) lebih mudah dan lebih cepat mempersiapkannya. Sedangkan kekurangannya adalah (1) hasilnya kurang dapat dipercaya, karena teknik penyekorannya subjektif, (2) penulis akan dapat menghindari kata-kata atau kalimat-kalimat tertentu yang dirasakannya sukar; dan (3) pemeriksaan hasil tes memerlukan waktu yang lama.

Metode tidak langsung adalah cara mengukur keterampilan menulis dengan mempergunakan tes bentuk objektif (misalnya bentuk pilihan berganda). Hasilnya dipergunakan untuk memperkirakan keterampilan menulis yang sebenarnya. Tes demikian disebut juga tes kemampuan dasar menulis (*writing ability*). Pengukuran metode langsung

dengan metode tidak langsung itu umumnya mempunyai korelasi yang tinggi.

Ada beberapa bentuk penilaian atau bentuk tugas kemampuan menulis bahasa Indonesia, yaitu (1) menyusun alinea, (2) menulis berdasarkan rangsangan visual, (3) menulis berdasarkan rangsangan suara, (4) menulis dengan rangsangan buku, (5) menulis laporan, (6) menulis surat, dan (7) menulis berdasarkan tema tertentu.

Meskipun penilaian kemampuan menulis yang lebih ideal adalah menyuruh murid untuk menulis secara esei, hal ini tidak berarti bentuk objektif tidak dapat dilakukan, melainkan dapat juga dilakukan dengan bentuk tugas menyusun alinea berdasarkan kalimat-kalimat yang telah disediakan. Berikut diberikan contoh penilaian objektif untuk menilai kemampuan dasar menulis dengan memperhatikan kata penghubung.

Bentuk-bentuk visual sebagai rangsangan untuk menghasilkan bahasa tertulis dapat berupa gambar atau film. Gambar yang memenuhi kriteria pragmatis untuk tugas atau menilai kemampuan menulis. Gambar sebagai rangsangan atau stimulus menilai atau tugas menulis baik diberikan kepada murid di kelas sekolah dasar atau bahasa target murid akan menghasilkan bahasa tulis walaupun masih sederhana. Kompleksitas gambar dapat bervariasi, bergantung kemampuan berbahasa murid yang diuji. Berikut dicontohkan tugas atau penilaian dengan rangsangan gambar.

Bentuk-bentuk suara yang dapat disajikan rangsangan tugas atau penilaian menulis dapat berupa suara langsung atau melalui media tertentu. Suara langsung adalah bentuk bahasa yang dihadirkan dalam komunikasi konkret seperti percakapan (*guneman*), diskusi, ceramah, dan sebagainya. Tugas atau penilaian yang dikerjakan murid adalah menulis karangan berdasarkan masalah yang dibicarakan dalam percakapan, diskusi, atau ceramah yang diikutinya. Tugas menulis dengan rangsangan suara ini memang bersifat tumpang tindih dengan tes kemampuan menyimak.

Bentuk suara yang tidak langsung dimaksudkan bahasa yang tidak langsung didengar dari orang yang menghasilkannya. Bentuk suara tersebut dapat dilakukan melalui rekaman radio dan televisi. Bentuk rangsangan dari radio atau televisi, sebaiknya kegiatan menyimak dan menulis karangannya dilakukan di rumah, sedangkan rekaman yang sudah disediakan di sekolah sebaiknya dilakukan di kelas saja.

Yang dimaksud dengan menulis dengan rangsangan buku adalah siswa distimulus dengan berbagai buku, karena buku sebagai bahan atau rangsangan untuk tugas menulis. Buku yang dijadikan perangsang tugas menulis dapat dibedakan menjadi dua macam, yaitu buku fiksi dan nonfiksi. Tugas menulis berdasarkan buku fiksi (*carita pondok, dongeng, novel, roman*) yang lebih cocok untuk dijadikan perangsang tugas menulis karangan. Tugas yang diberikan kepada murid cukup sederhana dengan cara menyusun kembali apa yang sudah dibaca dengan bahasa sendiri.

Untuk tingkatan pendidikan yang tinggi dapat dilakukan dengan tugas menulis resensi buku.

Seperti kita ketahui bahwa surat merupakan salah satu media komunikasi tertulis dalam kehidupan sehari-hari. Untuk itu, surat dapat dijadikan salah satu stimulus dalam melakukan penilaian menulis atau mengarang bahasa Indonesia. Tentu saja surat dalam berbahasa Indonesia yang sering digunakan adalah surat yang formal dan informal. Tugas yang diberikan kepada murid adalah menyuruh murid menulis surat pribadi atau surat yang lainnya.

Penilaian yang paling seriang dilakukan dalam mengukur kemampuan menulis kepada murid adalah dengan menyediakn tema-tema atau sejumlah tema yang dipilih atau berupa judul-judul yang harus dikembangkan oleh murid. Penyediaan berbagai tema yang akan dipilih akan memberikan kebebasan kepada murid untuk memberi judul karangannya dan mengembangkannya sesuai dengan minat dan kemampuan murid itu sendiri. Dalam bentuk penilaian ini guru atau penilai dapat memberikan petunjuk dua macam, yaitu pertama dengan memberi tema-tema yang dikembangkan oleh jurid dan kedua menentukan tema dengan kerangka karangannya yang dikembangkan oleh siswa sendiri.

Penilaian yang dilakukan terhadap karya (karangan) siswa biasanya bersifat holistik, impresif, dan selintas. Penilaian yang bersifat menyeluruh

berdasarkan kesan yang diperoleh dari membaca karangan secara selintas saja. Penilaian yang demikian jika dilakukan oleh orang yang ahli dan berpengalaman memang dapat dipertanggungjawabkan. Akan tetapi, keahlian itu belum tentu dimiliki oleh para guru di sekolah.

Berikut ini disajikan contoh-contoh model penilaian terhadap karangan siswa.

MODEL PENILAIAN TUGAS MENULIS DENGAN SKALA 10

No.	Aspek yang dinilai	Tingkatan skala 10										
1	Kualitas dan ruang lingkup	0	1	2	3	4	5	6	7	8	9	10
2	Organisasi dan penyajian	0	1	2	3	4	5	6	7	8	9	10
3	Gaya dan bentuk bahasa	0	1	2	3	4	5	6	7	8	9	10
4	Mekanik: tata bahasa, ejaan, kerapian tulisan	0	1	2	3	4	5	6	7	8	9	10
5	Respon afektif guru terhadap karangan	0	1	2	3	4	5	6	7	8	9	10

(Nurgiyantoro, 1988:304)

Selain model di atas, kita juga dapat memilih model pendekatan analitis yang lain, misalnya analisis unsur-unsur karangan seperti yang dikemukakan oleh Harris (196:(68-69) atau Halim (1974:100), yaitu unsur-unsur yang dinilai dalam kemampuan menulis adalah *content* (isi, gagasan yang dikemukakan), *form*

(organisasi isi), *grammar* (tata bahasa dan pola kalimat), *style* (gaya: pilihan struktur dan kosa kata, serta *mechanics* (ejaan). Berikut ini disajikan contoh model penilaian tugas menulis dengan pembobotan masing-masing-masing unsur kemampuan menulis.

MODEL PENILAIAN TUGAS MENULIS DENGAN PEMBOBOTAN
MASING-MASING UNSUR KEMAMPUAN MENULIS

No.	Unsur yang dinilai	Skor maksimum	Skor Siswa
1	Isi gagasan yang dikemukakan	35
2	Organisasi isi	25
3	Tata Bahasa	20
4	Gaya: pilihan struktur dan kosa kata	15
5	Ejaan	5

(lihat Nurgiyantoro, 1988:305)

Model penilaian kemampuan menulis yang ketiga dilakukan dengan rinci dan lebih teliti dalam pemberian skornya. Model ini diadopsi dari program ESL (*English as a Second Language*).

KRITERIA PENILAIAN KARANGAN

Nama Murid : _____
 Judul karangan : _____

Aspe	Skor	Kriteria
I S I	27 – 30	SANGAT BAIK - SEMPURNA: pada informai “subtansi” pengembangan tesis tuntas “relevan” dengan permasalahan dan tuntas.
	22 – 26	CUKUP – BAIK: informaai “substansi cukup” pengembangan tesis terbatas “relevan dengan masalah” tetapi tak lengkap.
	17 – 21	SEDANG- CUKUP: informasi terbatas “substansi kurang” pengembangan tesis tak cukup, permasalahan tak cukup.
	13 – 16	SANGAT KURANG: tak berisi “ tak ada substansi” tak ada pengembangan tesis’ tak ada permasalahan
O R G	18 – 20	SANGAT BAIK- SEMPURNA: ekspresi lancar, gagasan diungkapkan dengan jelas, “padat”, tertata dengan baik, urutan logis dan kohesif.

A N I S A S I	14 – 17	CUKUP – BAIK: ekspresi kurang lancar, kurang terorganisir, tetapi ide utama terlihat, bahan pendukung terbatas, urutan logis tetapi tak lengkap.
	10 - 13	SEDANG – CUKUP: ekspresi tak lancar, gagasan kacau, terpotong-potong, urutan dan pengembangan tak logis.
	7 - 9	SANGAT KURANG: tak komunikatif, tak terorganisir, tak layak nilai
K O S A K A T A	18 - 20	SANGAT BAIK – SEMPURNA: Pemanfaatan potensi kata canggih, pilihan kata dan ungkapan kata tepat, menguasai pembentukan kata
	14 – 17	CUKUP – BAIK: pemanfaatan potensi kata agak canggih, pilihan kata dan ungkapan kadang-kadang kurang tepat tetapi tak mengganggu.
	10 – 13	SEDANG – CUKUP: pemanfaatan potensi kata terbatas, sering terjadi kesalahan penggunaan kosa kata dan dapat merusak makna.
	7 - 9	SANGAT KURANG: pemanfaatan potensi kosa kata asal-asalan, pengetahuan tentang kosa kata rendah, tak layak nilai.
P	22 – 25	SANGAT BAIK - SEMPURNA: konstruksi kompleks tetapi efektif, hanya terjadi sedikit kesalahan penggunaan bentuk kebahasaan.

S	18 – 21	CUKUP – BAIK: konstruksi sederhana tetapi efektif, kesalahan kecil pada konstruksi kompleks, terjadi sejumlah kesalahan tetapi maknanya tidak kabur.
	11 – 17	SEDANG- CUKUP: terjadi kesalahan serius dalam konstruksi kalimat, makna membingungkan dan kabur.
	7 – 9	SANGAT KURANG: tak menguasai aturan sintaktis, terdapat banyak kesalahan, tak komunikatif, tak layak nilai.
M	5	SANGAT BAIK - SEMPURNA: menguasai aturan penulisan, hanya terdapat beberapa kesalahan ejaan.
	4	CUKUP – BAIK: kadang-kadang terjadi kesalahan ejaan tetapi tak mengaburkan makna.
	3	SEDANG- CUKUP: sering terjadi kesalahan ejaan, makna membingungkan atau kabur.
	2	SANGAT KURANG: tak menguasai aturan penulisan, terdapat banyak kesalahan ejaan, tulisan tak terbaca, tak layak nilai.

(lihat Hartfield, dkk., 1985:91 dan Nurgiyantoro, 1988:305-306).

d. Kemampuan Berbicara

Penilaian berbicara merupakan teknik pengukuran untuk mengumpulkan informasi mengenai kemampuan seseorang (siswa) dalam keterampilan berbicara. Informasi ini akan dipakai untuk menentukan nilai keterampilan berbicara.

Pada umumnya tes berbicara bukan hanya tes lisan melainkan juga tes perbuatan/penampilan, yakni tes nonverbal. Ini berarti yang dinilai bukan hanya perbuatan berbicara, melainkan juga proses/perbuatan dalam menghasilkan pembicaraan itu. Untuk itu, teknik tes berbicara dibantu oleh teknik observasi: penguji mengamati (bukan hanya mendengarkan) bagaimana peserta tes (testee) berbicara. Hal ini berlaku pada tes berbicara yang dilakukan secara langsung (*direct oral performance testing*).

Sebuah tes keterampilan terpadu, tes berbicara memadukan sejumlah komponen untuk dijadikan sasaran tes, yaitu (1) bahasa lisan yang digunakan, (2) isi pembicaraan, (3) teknik dan penampilan.

Teknik tes berbicara dapat digunakan dengan teknik bercakap-cakap, tanya jawab, wawancara, diskusi, debat, bermain peran, bercerita, berpidato, berceramah, laporan, dan teknik membacakan (membaca nyaring).

Ada beberapa bentuk penilaian berbicara bahasa Indonesia, yaitu (a) pembicaraan berdasarkan gambar, (b) wawancara, (c) bercerita, (d) pidato (*biantara*), dan diskusi.

Untuk mengungkap kemampuan berbicara bahasa Indonesia, gambar dapat dijadikan stimulus pembicaraan yang baik. Stimulus yang berupa gambar sangat baik dipergunakan untuk penilaian kemampuan berbicara murid-murid usia sekolah dasar. Akan tetapi, stimulus gambar pun dapat pula dipergunakan pada murid yang kemampuan berbahasanya lebih tinggi bergantung pada keadaan gambar yang dipergunakannya. Menurut Oller (1979: 47-8, 308-14) menyatakan bahwa gambar-gambar yang baik adalah gambar yang menarik siswa untuk mau berbicara atau mudah untuk mengungkapk kemampuan berbicara murid. Tugas-tugas yang diberikan kepada murid dapat berupa pemberian pertanyaan dan bercerita.

Ada beberapa cara untuk menilai tugas berpidato, Valette, 1977:149) mengembangkan teknik penilaian tugas-tugas laporan lisan dengan menggunakan skala 10. Berikut ini disajikan contoh model penilaian tugas berpidato (dan bercerita)

MODEL PENILAIAN TUGAS BERPIDATO .BERCERITA

No.	Aspek yang dinilai	Tingkatan skala
-----	--------------------	-----------------

1	Keakuratan informasi (sangat buruk – akurat sepenuhnya)	0	1	2	3	4	5	6	7	8	9	10
2	Hubungan antarinformaSI (sangat sedikit-berhubungan penuh)	0	1	2	3	4	5	6	7	8	9	10
3	Ketepatan struktur dan kosa kata (tidak tepat – tepat sekali)	0	1	2	3	4	5	6	7	8	9	10
4	Kelancaran (terbata-batalancer sekali)	0	1	2	3	4	5	6	7	8	9	10
5	Kewajaran unitan wacana (tak normal – normal)	0	1	2	3	4	5	6	7	8	9	10
6	Gaya pengucapan (kaku – wajar)	0	1	2	3	4	5	6	7	8	9	10
		Jumlah skor:										

4.3 Penilaian Berbasis Kelas (PBK)

4.1 Tujuan dan Fungsi Penilaian Berbasis Kelas

Penilaian Berbasis Kelas (PBK) merupakan suatu penilaian berdasarkan pada suatu proses pengumpulan, pelaporan dan penggunaan informasi tentang hasil belajar siswa yang diperoleh melalui pengukuran dengan menerapkan prinsip-prinsip penilaian, pelaksanaan berkelanjutan, bukti otentik, akurat dan konsisten sebagai akuntabilitas publik. PBK mengidentifikasi pencapaian kompetensi dan hasil belajar yang dikemukakan melalui pertanyaan yang jelas tentang standar yang harus dan telah dicapai disertai dengan peta kemajuan belajar siswa. PBK merupakan sebagian dari evaluasi dan merupakan komponen Kurikulum Berbasis Kompetensi.

Tujuan umum PBK adalah untuk memberikan penghargaan terhadap pencapaian belajar siswa dan memperbaiki program dan kegiatan pembelajaran. Secara khusus tujuan PBK adalah untuk memberikan (a) informasi tentang kemajuan belajar, (b) informasi yang dapat digunakan untuk membina kegiatan belajar lebih lanjut; (c) motivasi belajar siswa, dan melakukan bimbingan yang lebih tepat. PBK hendaknya menjamin bahwa hasil kerja siswa dan pencapaian belajarnya dapat diidentifikasi.

Fungsi PBK bagi siswa dan guru adalah untuk membantu (a) siswa dalam mewujudkan dirinya dengan mengubah atau mengembangkan perilakunya ke arah yang lebih baik dan maju; (b) siswa mendapat kepuasan atas apa yang dikerjakannya; (c) guru untuk menetapkan apakah metode mengajar yang digunakannya telah memadai atau tidak; dan (d) guru membuat pertimbangan dan keputusan administrasi.

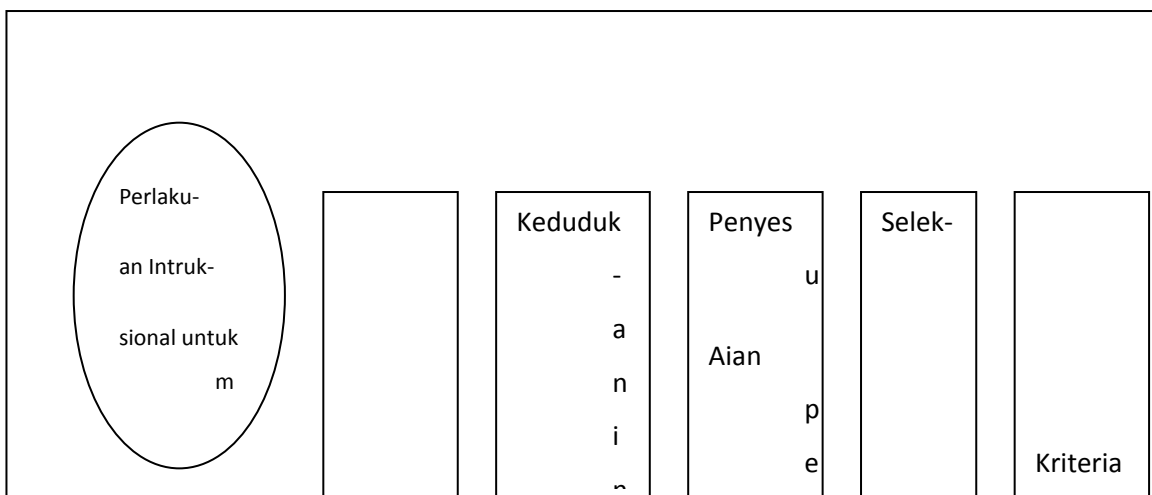
4.2 Prinsip-prinsip Penilaian Berbasis Kelas

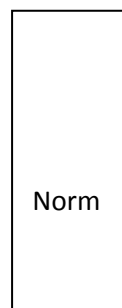
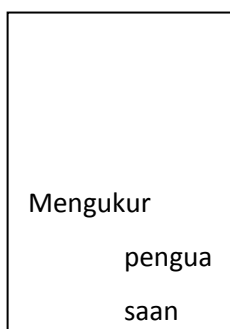
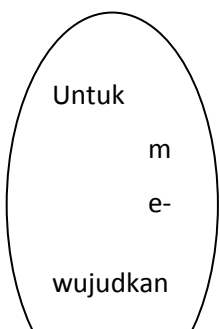
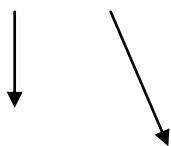
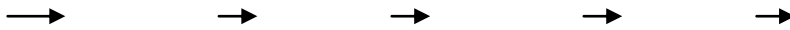
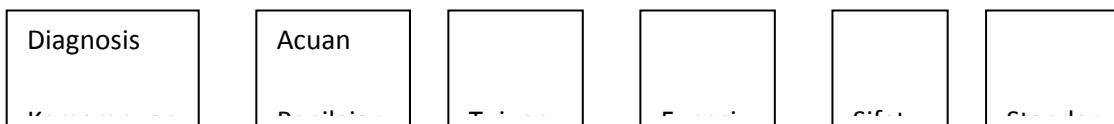
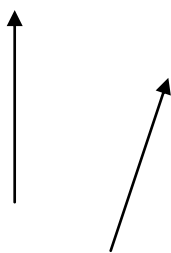
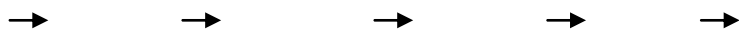
Prinsip-prinsip umum PBK adalah (a) valid, (b) mendidik, (c) berorientasi pada kompetensi, (d) adil dan objektif, (e) terbuka, (f) berkesinambungan, (g) menyeluruh, dan (h) bermakna. Pada segi lain ada dua prinsip khusus PBK: **Pertama**, apapun jenis penilaian harus memungkinkan adanya kesempatan yang terbaik bagi siswa untuk menunjukkan apa yang mereka ketahui dan pahami, serta mendemonstrasikan kemampuannya. Prinsip ini berimplikasi pada pelaksanaan PNK yang hendaknya dalam suasana yang bersahabat dan tidak mengancam, semua siswa mempunyai kesempatan dan mendapat perlakuan yang sama dalam menerima program pembelajaran sebelumnya dan selama proses PBK; siswa memahami secara jelas apa yang dimaksud dalam PBK, dan kriteria membuat keputusan atau hasil PBK hendaknya disepakati dengan siswa dan orang tua/wali.

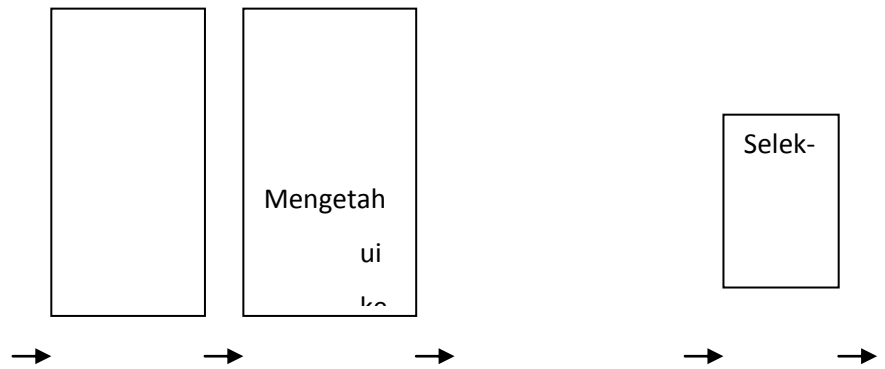
Kedua, setiap guru harus mampu melaksanakan prosedur PBK dan pencatatan secara tepat. Implikasi dari prinsip ini adalah bahwa prosedur PBK harus dapat diterima oleh guru dan dipahami secara jelas; prosedur PBK dan catatan hasil belajar siswa hendaknya mudah dilaksanakan sebagai bagian dari KBM dan tidak mengambil waktu yang berlebihan, catatan harian harus mudah dibuat, jelas, dan mudah dipahami, informasi yang diperoleh untuk menilai semua pencapaian belajar siswa dengan berbagai cara harus digunakan sebagaimana mestinya; penilaian pencapaian belajar siswa yang bersifat positif untuk pembelajaran selanjutnya; klasifikasi dan kesulitan belajar harus ditentukan sehingga mendapat bimbingan dan bantuan belajar yang wajar, hasil penilaian hendaknya menunjukkan kemajuan dan berkelanjutan bagi pencapaian belajar siswa; penilaian semua aspek yang berkaitan dengan pembelajaran yang efektif, peningkatan keahlian guru, dan pelaporan penampilan siswa kepada orang tua atau wali.

4.3 Acuan Penilaian Berbasis Kelas

Acuan yang digunakan dalam PBK dapat dilihat pada gambar berikut ini.







Gambar 1: Acuan Penilaian Berbasis Kelas

Keterangan: PAP = Penilaian Acuan Patokan

PAN = Penilaian Acuan Norma

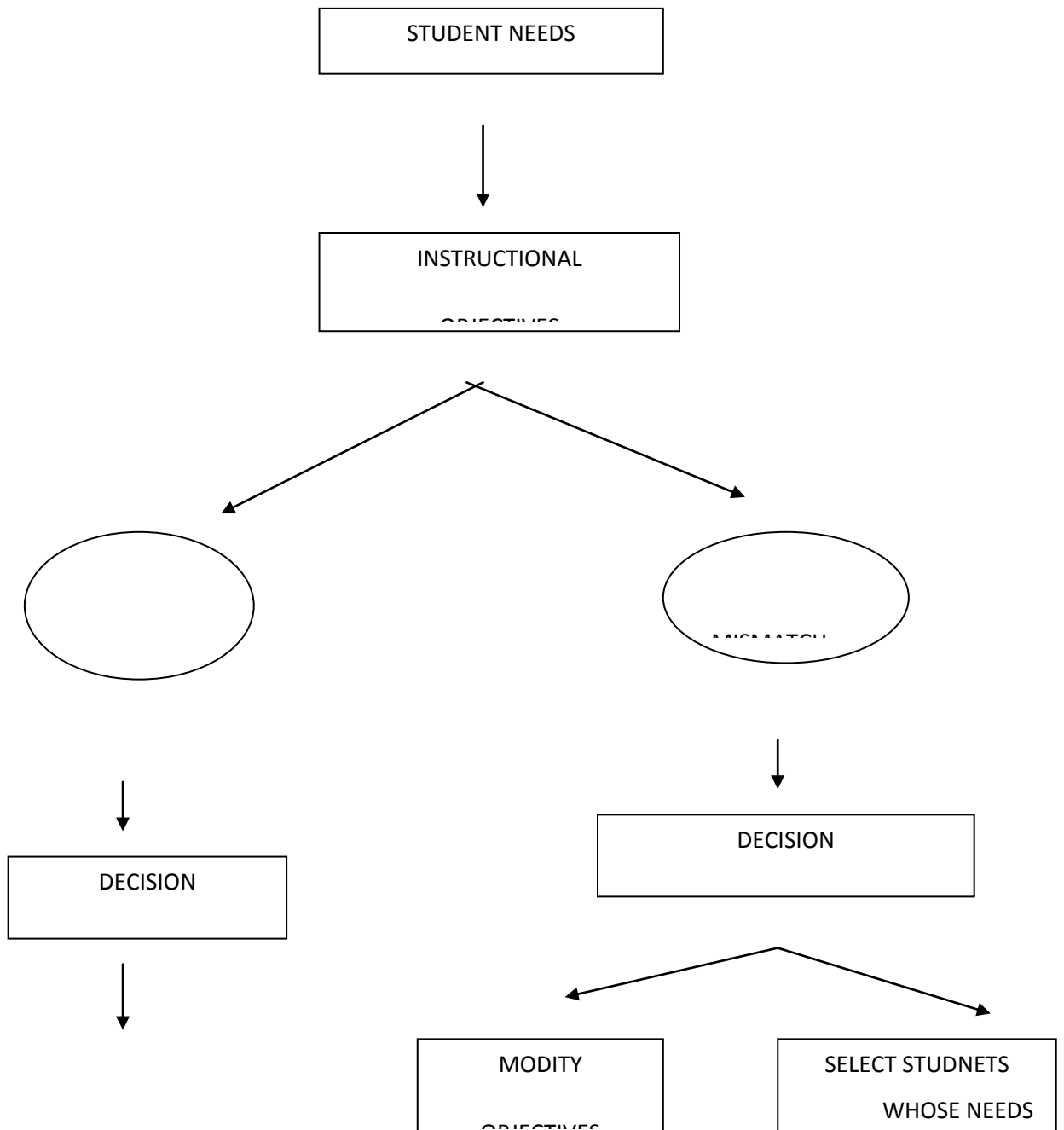
KD = Kompetensi Dasar

Dalam Kurikulum Berbasis Kompetensi siswa dituntut memiliki kemampuan dari hasil perbandingan antara pencapaian sebelum dan sesudah

pembelajaran dan kriteria penguasaan kompetensi yang ditentukan. Oleh sebab itu, dalam PBK lebih tepat apabila menggunakan penilaian acuan patokan (PAP).

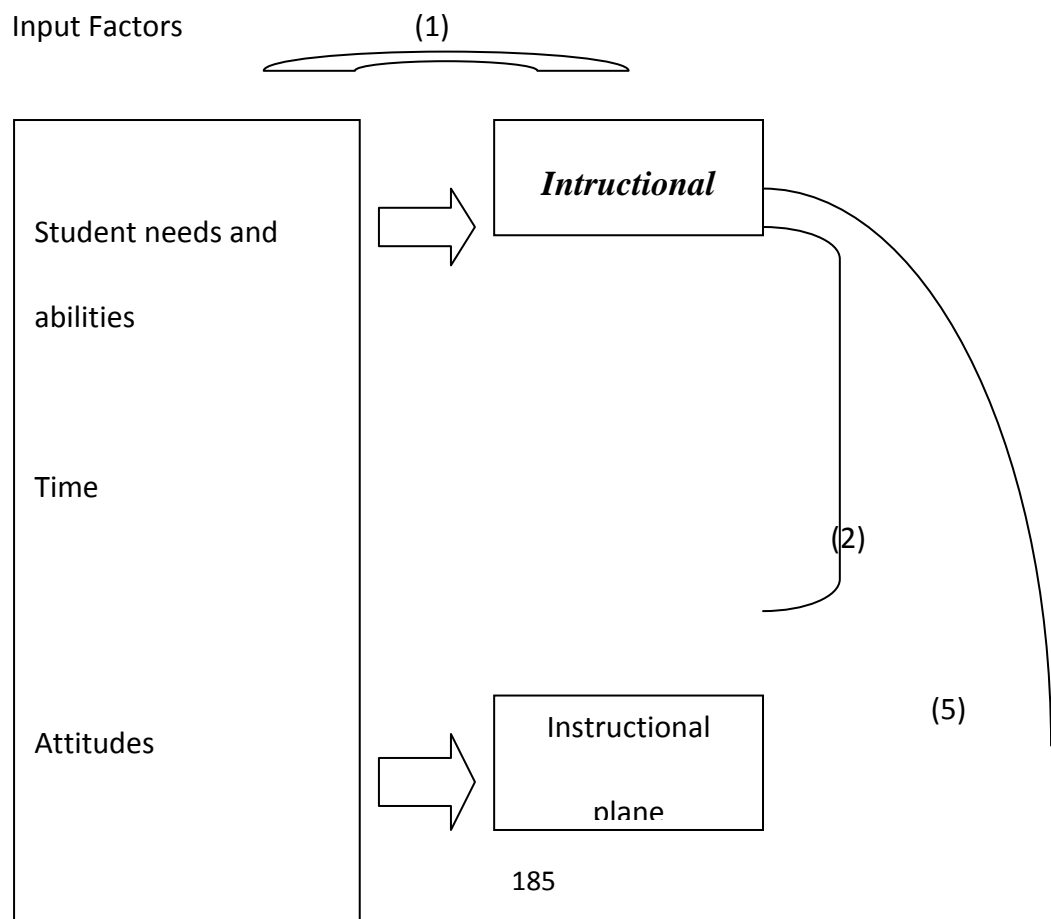
4.4 Strategi Penilaian Berbasis Kelas

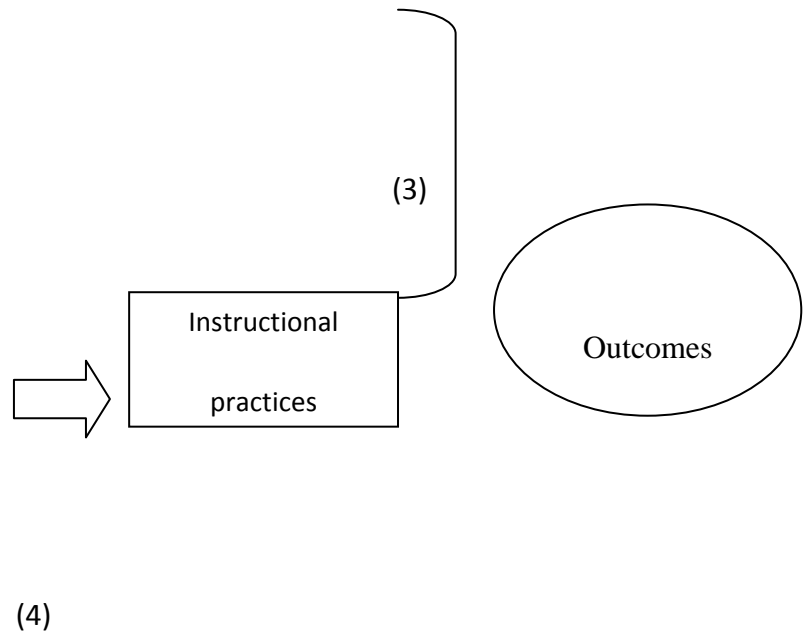
Di bawah ini disajikan dalam gambar 2 tentang bagaimana cara kita dalam mengambil suatu keputusan dalam PBK ((Richards, Ed.; 1998:41).



CONTINUE WITH
INSTRUCTIONAL

Strategi PBK terdiri atas lima langkah, yaitu dapat dilihat pada gambar di bawah ini.





(Richards, Ed.; 1998:43)

4.4 Kompetensi Berbahasa

Richard (1987:49) menjelaskan bahwa kompetensi komunikatif dalam

pembelajaran bahasa meliputi (1) pengetahuan mengenai gramatika dan

kosakata, (2) pengetahuan mengenai kaidah-kaidah berbicara, (3)

pengetahuan mengenai bagaimana cara menggunakan dan memberi

respons terhadap tindak-tutur, dan (4) pengetahuan mengenai

bagaimana cara menggunakan bahasa secara tepat dan memuaskan.

Selanjutnya, ia mengemukakan bahwa komponen-komponen

kompetensi komunikatif meliputi (1) kompetensi gramatikal, (2)

kompetensi sosiolinguistik, (3) kompetensi wacana, dan (4) kompetensi strategi.

Bachman (1990:87) menyebutnya bukan kompetensi komunikatif, tetapi kompetensi bahasa (*language competence*), yang meliputi (1) kompetensi organisasi dan kompetensi pragmatik. Kompetensi organisasi diklasifikasi lagi menjadi (a) kompetensi gramatikal dan (b) kompetensi wacana. Kompetensi pragmatik pun diklasifikasi menjadi (a) kompetensi *illocutionary competence* dan (b) kompetensi sosiolinguistik.

Kompetensi gramatikal menurut Bachman adalah kemampuan berbahasa dalam hal penguasaan dan penggunaan kaidah-kaidah bahasa, seperti kosakata, pembentukan kata, pembentukan kalimat, dan pembentukan bunyi/sistem penulisan. Sedangkan kompetensi wacana (*textual competence*) yaitu kemampuan siswa dalam penggunaan bahasa dalam aspek kekohesifan dan kekoherenan. *Illocutionary competence* mencakupi pemakaian bahasa yang berkaitan dengan fungsi-fungsi bahasa, seperti fungsi regulasi, fungsi heuristik, fungsi ideasional, fungsi imajinasi, fungsi personal, fungsi interpersonal, dan fungsi instrumental. Kompetensi sosiolinguistik meliputi kemahiran berbahasa dalam hal *sensitivity to dialect or variety, sensitivity to register, sensitivity to naturalness*, dan *references and figurative speech*.

Selanjutnya Littlewood (1981) mengemukakan bahwa ada dua jenis kompetensi komunikatif, yaitu pra komunikatif dan komunikatif. Yang dimaksud dengan kompetensi pra komunikatif adalah kemampuan berbahasa dalam aspek kompetensi struktural dan kuasi komunikatif; sedangkan kompetensi komunikatif meliputi kemampuan berbahasa dalam aspek komunikatif fungsional dan interaksi sosial.

Untuk dapat menyusun suatu tes Bahasa dan Sastra Indonesia yang baik dan terukur, pengembang tes perlu memperhatikan dan mempelajari dimensi-dimensi tes bahasa, yaitu (1) dimensi tujuan tes bahasa, (2) dimensi bentuk stimulus tes, (3) dimensi bentuk respons tes, (4) dimensi isi tes, (5) dimensi kemampuan teres, (6) dimensi teknik tes, dan (7) dimensi reliabilitas dan validitas tes.

Dimensi tujuan tes bahasa meliputi empat jenis, yaitu tes pencapaian atau tes kemajuan, tes sikap, tes diagnostik, dan tes penempatan.

Dimensi bentuk stimulus tes adalah satu runtunan stimulus dan respons. Oleh karena itu, dalam penyusunan tes bahasa perlu diperhatikan bentuk stimulus yang perlu dirancang oleh pengembang tes. Stimulus-stimulus tersebut dapat berupa lisan, tertulis, gambar-gambar, dan tindakan.

Dimensi bentuk respons merupakan jawaban yang diberikan oleh siswa dalam merespons stimulus yang diberikan oleh penguji. Respons yang diberikan oleh siswa dapat berupa lisan, tertulis, gambar-gambar, dan tindakan.

Dimensi isi tes bahasa berupa tes terpeggal dan tes terpadu. Tes terpeggal merupakan tes yang hendak mengukur kemampuan siswa dalam

menguasai ejaan dan tanda baca atau penguasaan kosa kata tertentu. Tes terpadu mengukur keseluruhan kemampuan siswa berbahasa sesuai dengan jenjang pendidikan dan tujuan pengajaran yang sudah ditetapkan.

Sesuai dengan tujuan pembelajaran Bahasa dan Sastra Indonesia yang menekankan aspek komunikatif, maka tes bahasa Indonesia pun harus mengukur kemampuan siswa dalam berbahasa untuk kepentingan komunikasi. Kemampuan komunikatif siswa yang dimaksudkan adalah kemampuan berbahasa Indonesia sesuai dengan situasi dan konteks.

4.5 Tes Bahasa Indonesia

4.6.1 Tes Struktur dan Kosakata

Struktur bahasa adalah unsur bahasa yang berupa kaidah yang mengatur susunan suatu bahasa; mulai dari bunyi, kata, kalimat, dan wacana. Respons yang diharapkan dalam tes struktur ini adalah penggunaan kaidah struktur bahasa Indonesia yang baik dan benar. Struktur bahasa yang harus dipelajari oleh peserta didik adalah bahasa yang normatif.

Tes struktur bahasa adalah teknik pengukuran untuk mengumpulkan informasi dalam rangka evaluasi penguasaan peserta didik terhadap struktur bahasa yang dipelajarinya. Tes struktur dapat dilakukan dengan pendekatan terpisah dan terpadu. Kedua pendekatan tes struktur ini

memiliki kekurangan dan kelebihan, akan tetapi sebaiknya dalam pengukuran struktur bahasa lebih baik keduanya digunakan.

Tes struktur bahasa Indonesia yang dilakukan di sekolah-sekolah (SMP-SMU) dewasa ini menggunakan pendekatan terpilah (*discrete-point approach test*), seperti yang kita ketahui dalam ulangan harian, ulangan umum, ujian sekolah, EBTA atau EBANAS.

Agar tes struktur bahasa Indonesia yang dilaksanakan di sekolah benar-benar dapat menggambarkan hasil belajar siswa seperti yang dituntut dalam kurikulum melalui pendekatan pengajaran bahasa yang komunikatif, maka seyogianya tes struktur bahasa Indonesia dilaksanakan sebagai berikut.

- a) Menggunakan kedua macam pendekatan tes bahasa (terpilah dan terpadu). Melalui pendekatan terpadu misalnya tes struktur bahasa dipadukan dengan salah satu keterampilan berbahasa yang produktif, seperti “tes penguasaan struktur bahasa melalui menulis atau mengarang”. Keuntungan yang kita peroleh ialah siswa dihadapkan pada perbuatan berbahasa yang aktual (otentik) juga bersifat komunikatif. Sedangkan dengan pendekatan terpilah, seperti yang lazim digunakan, yaitu melalui melengkapi kalimat dengan menggunakan bentuk kata, memilih kalimat yang benar, menafsirkan arti bentuk kata dan kalimat, menyusun kalimat yang susunannya kacau, menyempurnakan bentuk kata, dan mengubah bentuk kalimat.

- b) Ranah yang harus diutamakan dalam tes struktur adalah ranah aplikatif, analisis, sintesis, dan evaluasi.
- c) Pada tes struktur yang menggunakan pendekatan terpilah sebaiknya tidak hanya jenis pilihan berganda, akan tetapi menggunakan jenis isian dan esai terbatas.
- d) Sampel struktur bahasa yang dipilih sebagai bahan tes sebaiknya struktur yang dapat mewakili (representatif) struktur-struktur bahasa Indonesia yang lainnya.

Tes kosa kata adalah teknik pengukuran untuk mengumpulkan informasi dalam rangka mengevaluasi penguasaan kosa kata individu dalam suatu bahasa, baik kosa kata aktif maupun kosa kata pasif. Oleh karena kosa kata merupakan unsur bahasa yang terpilah (*separated object*), maka tesnya pun menggunakan pendekatan terpilah. Akan tetapi, tuntutan Kurikulum 1994, sebaiknya tes kosa kata bahasa Indonesia menggunakan pendekatan terpadu.

Ada dua masalah yang dihadapi dalam tes kosa kata, yaitu (1) dalam menentukan kosa kata yang akan diteskan atau respons kosa kata mana yang harus distimulus, (2) jumlah kosa kata yang tidak jelas, dan (3) menentukan jenis tes yang akan digunakan.

4.6.2 Tes Berbicara

Tes berbicara adalah teknik pengukuran untuk mengumpulkan informasi mengenai kemampuan seseorang (siswa) dalam keterampilan berbicara. Informasi ini akan dipakai untuk menentukan nilai keterampilan berbicara.

Pada umumnya tes berbicara bukan hanya tes lisan melainkan juga tes perbuatan/penampilan, yakni tes nonverbal. Ini berarti yang dinilai bukan hanya perbuatan berbicara, melainkan juga proses/perbuatan dalam menghasilkan pembicaraan itu. Untuk itu, teknik tes berbicara dibantu oleh teknik observasi: penguji mengamati (bukan hanya mendengarkan) bagaimana peserta tes (testee) berbicara. Hal ini berlaku pada tes berbicara yang dilakukan secara langsung (*direct oral performance testing*).

Sebuah tes keterampilan terpadu, tes berbicara memadukan sejumlah komponen untuk dijadikan sasaran tes, yaitu (1) bahasa lisan yang digunakan, (2) isi pembicaraan, (3) teknik dan penampilan.

Teknik tes berbicara dapat digunakan dengan teknik bercakap-cakap, tanya jawab, wawancara, diskusi, debat, bermain peran, bercerita, berpidato, berceramah, laporan, dan teknik membacakan (membaca nyaring).

4.6.3 Tes Menulis/Mengarang

Ada dua metode yang sering digunakan dalam pengukuran kemampuan menulis atau mengarang, yaitu metode langsung dan metode tidak langsung (Halim, 1982:115-116). Metode langsung merupakan tes keterampilan menulis langsung dilaksanakan dengan cara pelaksana tes (guru) langsung menyuruh siswa atau peserta tes menulis atau mengarang topik-topik atau judul-judul karangan tertentu. Keunggulan metode langsung adalah (a) dapat mengukur kemampuan tertentu (kemampuan menyusun, menghubungkan serta memakai bahasa yang dikarangnya dapat lebih efektif, (b) mempunyai potensi untuk mendorong peserta mengerjakan tugasnya sebaik-baiknya; dan (c) lebih mudah dan lebih cepat mempersiapkannya. Sedangkan kekurangannya adalah (a) hasilnya kurang dapat dipercaya, karena teknik penyekorannya subjektif, (b) penulis akan dapat menghindari kata-kata atau kalimat-kalimat tertentu yang dirasakannya sukar; dan (c) pemeriksaan hasil tes memerlukan waktu yang lama.

Metode tidak langsung adalah cara mengukur keterampilan menulis dengan mempergunakan tes bentuk objektif (misalnya bentuk pilihan berganda). Hasilnya dipergunakan untuk memperkirakan keterampilan menulis yang sebenarnya. Tes demikian disebut juga tes kemampuan dasar menulis (*writing ability*).

Menurut temuan beberapa peneliti, seperti Breland, Colon & Bogosa (1976), Breland & Gayner (1979), dan Moss, Cols (1981) ternyata hasil pengukuran metode langsung dengan metode tidak langsung itu umumnya mempunyai korelasi yang tinggi (Stinggins: 1982:347).

Richard (1987:49) menjelaskan bahwa kompetensi komunikatif dalam pembelajaran bahasa meliputi (1) pengetahuan mengenai gramatika dan kosakata, (2) pengetahuan mengenai kaidah-kaidah berbicara, (3) pengetahuan mengenai bagaimana cara menggunakan dan memberi respons terhadap tindak-tutur, dan (4) pengetahuan mengenai bagaimana cara menggunakan bahasa secara tepat dan memuaskan. Selanjutnya, ia mengemukakan bahwa komponen-komponen kompetensi komunikatif meliputi (1) kompetensi gramatikal, (2) kompetensi sosiolinguistik, (3) kompetensi wacana, dan (4) kompetensi strategi.

Bachman (1990:87) menyebutnya bukan kompetensi komunikatif, tetapi kompetensi bahasa (*language competence*), yang meliputi (1) kompetensi organisasi dan kompetensi pragmatik. Kompetensi organisasi diklasifikasi lagi menjadi (a) kompetensi gramatikal dan (b) kompetensi wacana. Kompetensi pragmatik pun diklasifikasi menjadi (a) kompetensi *illocutionary competence* dan (b) kompetensi sosiolinguistik.

Kompetensi gramatikal menurut Bachman adalah kemampuan berbahasa dalam hal penguasaan dan penggunaan kaidah-kaidah bahasa, seperti kosakata, pembentukan kata, pembentukan kalimat, dan pembentukan bunyi/sistem penulisan. Sedangkan kompetensi wacana (*textual competence*) yaitu kemampuan siswa dalam penggunaan bahasa dalam aspek kekohesifan dan kekoherenan. *Illocutionary competence* mencakupi pemakaian bahasa yang berkaitan dengan fungsi-fungsi bahasa, seperti fungsi regulasi, fungsi heuristik, fungsi ideasional, fungsi imajinasi, fungsi personal, fungsi interpersonal, dan fungsi instrumental. Kompetensi sosiolinguistik meliputi kemahiran berbahasa dalam hal *sensitivity to dialect or variety, sensitivity to register, sensitivity to naturalness, dan references and figures of speech.*

Selanjutnya Littlewood (1981) mengemukakan bahwa ada dua jenis kompetensi komunikatif, yaitu pra komunikatif dan komunikatif. Yang dimaksud dengan kompetensi pra komunikatif adalah kemampuan berbahasa dalam aspek kompetensi struktural dan kuasi komunikatif. Sedangkan kompetensi komunikatif meliputi kemampuan berbahasa dalam aspek komunikatif fungsional dan interaksi sosial.

Untuk dapat menyusun suatu tes bahasa Indonesia yang baik dan terukur, pengembang tes perlu memperhatikan dan mempelajari dimensi-dimensi tes bahasa, yaitu (1) dimensi tujuan tes bahasa itu, (2) dimensi bentuk stimulus tes,

(3) dimensi bentuk respons tes, (4) dimensi isi tes, (5) dimensi kemampuan teres, (6) dimensi teknik tes, dan (7) dimensi reliabilitas dan validitas tes.

Dimensi tujuan tes bahasa meliputi empat jenis, yaitu tes pencapaian atau tes kemajuan, tes sikap, tes diagnostik, dan tes penempatan.

Dimensi bentuk stimulus tes adalah satu runtunan stimulus dan respons. Oleh karena itu, dalam penyusunan tes bahasa perlu memperhatikan bentuk stimulus yang perlu dirancang oleh pengembang tes. Stimulus-stimulus tersebut dapat berupa lisan, tertulis, gambar-gambar, dan tindakan.

Dimensi bentuk respons merupakan jawaban yang diberikan oleh siswa dalam merespons stimulus yang diberikan oleh penguji. Respons yang diberikan oleh siswa dapat berupa lisan, tertulis, gambar-gambar, dan tindakan.

Dimensi isi tes bahasa berupa tes terpecah dan tes terpadu. Tes terpecah merupakan tes yang hendak mengukur kemampuan siswa dalam penguasaan ejaan dan tanda baca atau penguasaan kosa kata tertentu. Tes terpadu mengukur keseluruhan kemampuan siswa berbahasa sesuai dengan jenjang pendidikan dan tujuan pengajaran yang sudah ditetapkan.

Sesuai dengan tujuan pembelajaran bahasa Indonesia yang menekankan aspek komunikatif, maka tes bahasa Indonesia pun harus mengukur kemampuan siswa dalam berbahasa untuk kepentingan komunikasi. Kemampuan komunikatif siswa yang dimaksudkan adalah kemampuan berbahasa Indonesia sesuai dengan situasi dan konteks. Indikator kemampuan yang dites itu adalah tingkat kompetensi komunikatif.

Dimensi teknik tes bahasa meliputi dikte, esei (mengarang), wawancara, pilihan berganda, tes rumpang (*cloze test*), dan terjemahan.

Tes bahasa yang baik adalah tes yang memiliki reliabilitas dan validitas yang tinggi.

