

# **EVALUASI DALAM PEMBELAJARAN BAHASA**

**DIKTAT**  
**Mata Kuliah Tes dan Penilaian Bahasa**  
**Untuk Mahasiswa Program Studi**  
**Pendidikan Bahasa Inggris**  
**FITK IAIN SURAKARTA**



**Oleh:**

**Dr. Imroatus Solikhah, M.Pd**

**PROGRAM STUDI PENDIDIKAN BAHASA INGGRIS**  
**FAKULTAS ILMU TARBIYAH DAN KEGURUAN**  
**INSTITUT AGAMA ISLAM NEGERI**  
**SURAKARTA**  
**Januari, 2017**

## LEMBAR PENGESAHAN

Judul : **Evaluasi dalam Pembelajaran Bahasa**  
Jenis : DIKTAT Mata Kuliah  
Penulis : Dr. Imroatus Solikhah, M.Pd  
NIP : 19770316 200912 2002  
Unit Kerja : Program Pendidikan Bahasa Inggris  
Fakultas Ilmu Tarbiyah dan Keguruan  
IAIN Surakarta  
Waktu Penulisan : Januari 2017

Surakarta, Pebruari 2017

Mengetahui:  
Dekan FITK,

Penulis,

**Dr. H. Giyoto, M.Hum**  
**NIP. 1967 7022 4199 031 003**

**Dr. Imroatus Solikhah, M.Pd**  
**NIP. 19770316 200912 2002**

## KATA PENGANTAR

Alhamdulillah. Segala puji bagi Allah seru sekalian alam yang telah melimpahkan karunia dan nikmat yang tidak ternilai harganya. Diktat, *Evaluasi dalam Pembelajaran Bahasa*, ini disusun untuk memenuhi kebutuhan sendiri dalam mata kuliah Evaluasi Pembelajaran Bahasa di Program Studi Pendidikan Bahasa Inggris Fakultas Ilmu Tarbiyah dan Keguruan tahun 2015. Naskah awal Diktat ini dikumpulkan berdasarkan pengalaman mengajar mata kuliah Evaluasi Pembelajaran Bahasa pada tahun-tahun sebelumnya. Susunan materi dan urutan tingkat kesulitan dalam Diktat telah dirancang cukup cermat berdasarkan pengalaman dalam kelas sebagai uji coba materi.

Materi Diktat ini dikembangkan mulai dari konsep tes dan evaluasi bahasa, syarat tes yang baik, interpretasi hasil tes, validitas, reliabilitas, tes uraian, dan contoh pengembangan tes membaca dan tes menulis. Perkembangan tes terbaru juga dimasukkan sebagai tema tersendiri yang dikemas dalam penilaian otentik. Dengan kerangka penyajian seperti itu diharapkan mahasiswa bisa memperoleh pengetahuan teoritis mengenai evaluasi dalam pembelajaran bahasa secara memadai dan bisa mengembangkan tes sendiri dengan mengacu pada contoh proses dan pengembangannya. Secara substantif, materi dalam Diktat ini bisa dianggap cukup memadai untuk pembekalan mahasiswa Program Sarjana dalam melaksanakan Praktik Mengajar di lapangan, pengembangan instrumen untuk penelitian skripsi, dan pengembangan alat evaluasi secara mandiri di luar perkuliahan.

Semoga Allah memberikan berkah atas jerih payah yang penulis lakukan dan para mahasiswa, kolega, atau pembaca pada umumnya melalui kritik dan saran untuk perbaikan Diktat ini.

Surakarta, Januari 2017  
Penulis,

**Dr. Imroatus Solikhah, M.Pd**

## DAFTAR ISI

<b>LEMBAR PENGESAHAN</b>	<b>ii</b>
<b>KATA PENGANTAR</b>	<b>iii</b>
<b>DAFTAR ISI</b>	<b>iv</b>
<b>BAB 1 TES, PENILAIAN, DAN PENGAJARAN</b>	
PENDAHULUAN	1
DEFINISI TES	2
PENILAIAN, PENGUKURAN, EVALUASI DAN PENGUJIAN	4
Penilaian Informal dan Penilaian Formal	8
Penilaian Formatif dan Sumatif	9
Tes Kriteria Norma dan Kriteria Patokan	10
PENDEKATAN DALAM TES BAHASA: SEJARAH SINGKAT	10
Tes Diskrit dan Integratif	11
Tes Bahasa Komunikatif	13
Penilaian Berbasis Kinerja	14
ISU-ISU MUTAKHIR DALAM TES BAHASA	15
Pandangan Baru dalam Intelegensi	16
Penilaian Tradisional dan Alternatif	18
Penilaian Berbasis Komputer	18
PERMASALAHAN	20
PEMBAHASAN	20
Model Penilaian Bahasa di Sekolah	21
Sistem Penilaian di Indonesia	23
Penguasaan Guru Terhadap Model Penilaian	25
SIMPULAN	27
DAFTAR RUJUKAN	28
<b>BAB 2 FUNGSI EVALUASI DALAM PEMBELAJARAN</b>	
PENDAHULUAN	30
PERKEMBANGAN ALIRAN DALAM TES	32
PENDEKATAN DALAM TES	33
PERMASALAHAN DALAM EVALUASI	37
Jenis Tes	37
Format Tes	39
Syarat Tes Yang Baik	40
Interpretasi Hasil Tes	42
Authentic Assessment	44

PENUTUP	45
DAFTAR RUJUKAN	46
<b>BAB 3 TES BUATAN GURU DAN TES TERSTANDAR</b>	
DEFINISI	47
CIRI-CIRI TES YANG BAIK	48
Validitas	48
Validitas Empiris	50
Validitas Logis	51
Reliabilitas	52
Analisis Hasil Tes	54
Tingkat Kesulitan (p)	55
Daya Pembeda (D)	55
<b>BAB 4 PENGEMBANGAN TES URAIAN</b>	
LATAR BELAKANG	57
TAHAPAN PENYUSUNAN BUTIR SOAL	58
CONTOH BUTIR PERTANYAAN	58
CONTOH TEKS SOAL	60
<b>BAB 5 CONTOH PENGEMBANGAN TES MEMBACA</b>	
LATAR BELAKANG	65
PERMASALAHAN	67
LANDASAN TEORI MEMBACA PEMAHAMAN	67
Membaca Literal	68
Jenis Pertanyaan dalam Membaca Pemahaman	69
KERANGKA PENGEMBANGAN	71
Definisi Konseptual	72
Definisi Operasional	73
Kisi-Kisi Soal	73
Proses Ujicoba	73
Teknik Analisis	75
<i>Teknik Manual</i>	75
<i>Analisis Menggunakan SPSS Release 11</i>	77
HASIL UJI COBA	78
Hasil Analisis Teknik Manual	78
<i>Tingkat Kesulitan Butir Soal</i>	78
<i>Hasil Analisis Daya Pembeda</i>	80
<i>Analisis Tingkat Kesulitan dan Daya Beda</i>	82
<i>Hasil Analisis Distraktor</i>	83
<i>Hasil Analisis Validitas</i>	85
<i>Hasil Analisis Reliabilitas</i>	86
<i>Analisis Reliabilitas Teknik Test-Retest</i>	86

Hasil Analisis SPSS Release 11	88
PEMBAHASAN	92
SIMPULAN	93
DAFTAR RUJUKAN	94
LAMPIRAN BUTIR SOAL READING	96
<b>BAB 6 CONTOH PENGEMBANGAN TES MENULIS ESAI</b>	
KERANGKA TEORITIS	105
DEFINISI KONSEPTUAL	110
DEFINISI OPERASIONAL	110
KISI-KISI INSTRUMEN	111
BUTIR-BUTIR INSTRUMEN ANGKET	111
DAFTAR RUJUKAN	116

# **BAB 1**

## **TES, PENILAIAN, DAN PENGAJARAN**

### **PENDAHULUAN**

Begitu pentingnya suatu tes, sampai-sampai ada ungkapan, kita hidup karena tes dan bahkan (kadang) meninggal juga karena tes. Kata tes umumnya berkonotasi tidak positif dan kurang menyenangkan karena kata tes menimbulkan rasa cemas dan kurang percaya diri. Namun tes adalah sesuatu yang tidak bisa dihindari sebagaimana matahari terbit dari arah timur setiap hari. Kata tes umumnya dianggap memiliki kemiripan makna dengan penilaian, pengukuran, dan evaluasi yang diperoleh melalui pengujian mengiringi pembelajaran. Secara umum, istilah penilaian digunakan lebih luas untuk menunjukkan proses pengambilan informasi hasil belajar, pengamatan, atau hasil kajian yang dilakukan secara formal dan informal.

Dalam pengajaran, penilaian hasil belajar kini tidak lagi cukup jika hanya didasarkan penilaian kognitif saja. Stronge (2006:101) mengemukakan proses belajar mengajar sangatlah kompleks dan guru mengalami persoalan yang tidak menentu untuk menetapkan balikan dari siswa. Karena itu, penilaian kelas harus melibatkan berbagai hal yang bisa memberikan informasi yang bertanggungjawab untuk penyelenggaraan pendidikan.

Sejalan dengan paradigma pembelajaran konstruktivis, penilaian kognitif kini selalu dibarengi dengan penilaian afektif dan psikomotorik. Menurut Gronlund (1985:34) untuk menilai keberhasilan tujuan pembelajaran ialah dengan menggunakan pedoman taksonomi tujuan pembelajaran yang dikemukakan Bloom dkk (1956). Taksonomi ini membagi tujuan pembelajaran ke dalam tiga domain: (1) kognitif yang menguji pengetahuan dan kemampuan intelektual, (2) afektif, yang menunjukkan sikap, minat, rasa menghargai, dan pengambilan keputusan, dan (3) psikomotorik, yang menunjukkan kemampuan motorik atau gerak.

Dengan perkembangan model pembelajaran inovatif dan konsep kecerdasan majemuk (multiple intelligence) taksonomi tujuan pembelajaran semakin penting maknanya dan digunakan sebagai landasan penilaian secara luas. Penilaian yang sebelumnya berbasis pada tes (paper-pencil test) kini berkembang menjadi penilaian autentik atau penilaian berbasis kelas. Penilaian berbasis kelas, hakikatnya ialah penilaian otentik yang didasarkan pada kenyataan kegiatan belajar di kelas senyatanya (Richards, 1996; Stronge, 2006). Penilaian otentik melibatkan selain tes, juga portofolio, proyek, kinerja, dan menulis lanjut (Nitko & Brookhart, 2007; Johnson, 1996).

Artikel ini dikembangkan berdasarkan isi bab I dari buku tulisan H. Douglas Brown (2004), *Language Assessment: Principles and Classroom Practices*. Uraian dari berbagai sumber digunakan untuk melengkapi kajian dan perspektif lebih lengkap. Komentar dan ulasan mengenai hal-hal yang secara empiris terkait dengan penggunaan dan perkembangan penilaian di Indonesia juga diberikan untuk melengkapi.

## **DEFINISI TES**

Ada lima istilah yang terkait dengan konsep penilaian yang digunakan untuk mengetahui keberhasilan belajar peserta didik, yaitu tes, pengukuran, pengujian, penilaian, dan evaluasi. Dalam makna sederhana, tes ialah metode untuk mengukur kemampuan seseorang, pengetahuan atau kinerja dalam bidang tertentu. Gronlund (1985:4) mendefinisikan tes sebagai alat atau prosedur sistematis untuk mengukur perilaku subjek. Tes bertujuan menjawab pertanyaan sejauh mana subjek mampu menunjukkan sesuatu, baik jika dibandingkan dengan individu lain atau dengan kriteria tertentu. Definisi tersebut memiliki empat makna yang berbeda substansinya.

**Pertama**, tes sebagai metode dan instrumen, terdiri dari seperangkat teknik, prosedur atau item yang menuntut kinerja tertentu dari peserta yang menjawab tes. Agar tes berkualitas, metode yang digunakan haruslah eksplisit dan terstruktur: pilihan ganda dengan pertanyaan dan jawaban yang sudah ditentukan dengan jelas;

interview lisan berdasarkan daftar pertanyaan yang sudah disusun dan daftar cek yang diharapkan diisi sesuai dengan jawaban yang diharapkan.

**Kedua**, tes harus mengukur. Ada tes yang mengukur kemampuan umum dan ada juga tes yang mengukur kompetensi atau tujuan khusus. Tes profisiensi dengan keterampilan majemuk mengukur tingkat kemampuan umum; quiz untuk mengetahui penggunaan yang benar artikel mengukur pengetahuan khusus. Bagaimana hasil atau pengukuran digunakan memiliki variasi. Penilaian berbasis kelas menggunakan esai jawaban pendek dinilai menggunakan huruf atau komentar. Tes yang lebih luas pesertanya seperti tes terstandar dinilai menggunakan skor, persentil atau angka-angka yang lebih rinci. Jika suatu instrumen tidak menunjukkan bentuk spesifik dalam menjelaskan hasil, tidak bisa disebut tes.

**Ketiga**, tes mengukur kemampuan individu, pengetahuan atau kinerja. Pemberi tes perlu mengetahui siapa yang dites. Apa saja pengalaman yang dimiliki peserta tes sebelumnya dan apa latar belakangnya? Apakah tes tersebut sesuai dengan kemampuan peserta tes? Bagaimana pemberi tes menginterpretasi skor yang diperoleh peserta tes?

Tes mengukur kinerja, tetapi hasilnya menunjukkan kemampuan peserta tes atau menggunakan istilah dalam linguistik disebut kompetensi. Umumnya tes bahasa mengukur kemampuan seseorang dalam berbahasa, yaitu, berbicara, menulis, membaca, atau menyimak. Tidak biasa ditemukan suatu tes dirancang untuk mengukur pengetahuan tentang bahasa dari peserta tes: mendefinisikan kosakata, menjelaskan aturan gramatika, atau mengidentifikasi fitur retorika suatu wacana. Tes berbasis kinerja membuat sampel kemampuan penggunaan bahasa secara nyata dari peserta tes, tetapi pemberi tes menginfer kemampuan umum berdasarkan sampel tersebut. Tes membaca pemahaman hanya terdiri dari beberapa teks pendek diikuti dengan beberapa pertanyaan. Berdasarkan hasil jawaban pertanyaan tersebut, penguji menyimpulkan kemampuan tertentu peserta tes dalam membaca.

**Keempat**, tes mengukur domain tertentu. Tes profisiensi misalnya, walaupun tesnya hanya mengukur sampel beberapa skill, sebenarnya domain tes tersebut

menunjukkan kemampuan bahasa secara menyeluruh. Jenis tes yang lain bisa lebih khusus lagi. Tes *pronunciation* misalnya, bisa hanya mengukur pasangan minimal suatu fonem. Tes yang disusun dengan baik merupakan instrumen yang memberikan pengukuran akurat dari kemampuan peserta tes dalam domain tertentu. Definisi tersebut nampak sederhana tetapi kenyataannya, menyusun tes yang baik merupakan pekerjaan kompleks yang melibatkan ilmu dan seni.

### **PENILAIAN, PENGUKURAN, EVALUASI DAN PENGUJIAN**

Kata *assessment* atau penilaian sering disamakan dengan testing, walaupun sebenarnya menilai dan mengetes (testing) tidaklah sama. Kedua kata tersebut tidak sinonim. Tes disiapkan dalam bentuk prosedur administratif yang terjadi pada waktu yang bisa diidentifikasi dalam kurikulum ketika pembelajar menyelesaikan seluruh materi yang diajarkan dan bisa disiapkan dalam kinerja terbaik, dan pembelajar mengetahui bahwa jawabannya akan diukur dan dievaluasi.

Penilaian merupakan rangkaian kegiatan untuk memperoleh, menganalisis, dan menafsirkan data tentang proses dan hasil belajar peserta didik yang dilakukan secara sistematis dan berkesinambungan, sehingga menjadi informasi yang bermakna dalam pengambilan keputusan. Jadi, penilaian adalah proses merefleksikan data untuk membuat suatu keputusan (Campbell, 1998). Penilaian merupakan proses pengumpulan, penganalisisan, dan penafsiran informasi secara sistematis untuk menentukan seberapa jauh tujuan telah tercapai. Menilai juga berarti suatu proses untuk memberi makna terhadap suatu gejala berdasarkan kriteria tertentu (Harsiati, 2001).

Penilaian menunjukkan proses yang sedang berjalan yang meliputi domain yang lebih luas. Ketika seorang siswa menjawab sebuah pertanyaan, memberikan komentar, atau mencoba mengemukakan kata atau struktur baru, guru secara tidak sadar menilai kinerja siswa tersebut. Hasil pekerjaan tertulis merupakan kinerja yang hakikatnya dinilai sendiri, dinilai guru atau orang lain. Aktivitas membaca dan menyimak selalu memerlukan kinerja produktif yang secara implisit dinilai oleh guru

walaupun secara sederhana. Guru yang baik tidak pernah berhenti menilai siswa baik secara kebetulan atau disengaja.

Dengan demikian, tes merupakan bagian dari penilaian; tes bukanlah sekadar bentuk penilaian yang bisa dibuat guru. Tes bisa menjadi alat yang bermanfaat, tetapi tes hanyalah satu dari sekian banyak prosedur dan tugas yang bisa digunakan guru untuk menilai.

Masalahnya sekarang, andaikan seseorang melakukan penilaian pada setiap kali mengajar suatu topik, apakah semua pengajaran melibatkan penilaian? Apakah guru secara konsisten membuat penilaian walaupun tanpa ada interaksi? Jawabnya tergantung pada perspektif mana yang digunakan. Untuk pembelajaran yang dilakukan secara optimal, siswa dalam kelas harus memiliki kebebasan mencoba, mencoba hipotesisnya sendiri tanpa harus merasa bahwa mereka dinilai kesalahannya dan upaya coba-coba. Dengan cara yang sama dalam turnamen tennis, pemain harus memiliki kebebasan melatih kemampuannya tanpa implikasi terhadap hasil akhir peringkatnya. Jadi, siswa juga harus memiliki kebebasan mempraktikkan kemampuan berbahasa tanpa harus secara formal dinilai. Pengajaran (teaching) memberikan latihan permainan dalam mempelajari bahasa: kesempatan bagi pembelajar untuk mendengarkan, berpikir, mengambil resiko, menentukan tujuan, dan memproses balikan dari guru dan kemudian mendaur ulang melalui kemampuan yang mereka coba kuasai.

Dalam waktu yang sama, selama aktivitas latihan berlangsung, guru sebenarnya mengamati kinerja siswa dan membuat berbagai evaluasi terhadap setiap pembelajar. Dalam kelas yang ideal, pengamatan memberi petunjuk bagi guru untuk memberikan pembelajaran pada setiap siswa.

Selama mengadakan penilaian, guru juga melakukan pengukuran (measurement), yaitu proses penetapan ukuran terhadap suatu gejala menurut aturan tertentu (Guilford, 1982). Jadi, pengukuran merupakan proses untuk mengkuantifikasi sejauh mana seseorang atau sesuatu memiliki bobot tertentu. Contoh pengukuran antara lain menentukan kualitas atau karakteristik suatu hal. Pengukuran memberikan data lebih

objektif dan bisa digunakan untuk membuat perbandingan secara lebih jelas. Misalnya, pernyataan: “*Gedung tinggi dan besar,*” akan lebih jelas jika diubah menjadi: “*Gedung 3 lantai berukuran 100x100m*” (Gay, 1985:8).

Pengukuran digunakan untuk melihat unjuk kerja atau kemampuan peserta didik dengan menggunakan suatu standar. Pengukuran dapat menggunakan tes dan non-tes. Pengukuran pendidikan bisa bersifat kuantitatif atau kualitatif. Kuantitatif hasilnya berupa angka, sedangkan kualitatif hasilnya bukan angka (berupa predikat atau pernyataan kualitatif, misalnya sangat baik, baik, cukup, kurang, sangat kurang), disertai deskripsi penjelasan prestasi peserta didik. Pengujian merupakan bagian dari pengukuran yang dilanjutkan dengan kegiatan penilaian.

Penilaian sebagai istilah umum untuk penilaian mencakup semua metode yang biasa digunakan untuk menilai unjuk kerja individu atau kelompok peserta didik. Proses penilaian mencakup pengumpulan bukti yang menunjukkan pencapaian belajar peserta didik. Penilaian merupakan suatu pernyataan berdasarkan sejumlah fakta untuk menjelaskan karakteristik seseorang atau sesuatu (Griffin & Nix, 1991). Penilaian mencakup semua proses pembelajaran. Oleh karena itu, kegiatan penilaian tidak terbatas pada karakteristik peserta didik saja, tetapi juga mencakup karakteristik metode mengajar, kurikulum, fasilitas, dan administrasi sekolah.

Instrumen penilaian untuk peserta didik dapat berupa metode dan/atau prosedur formal atau informal untuk menghasilkan informasi tentang peserta didik. Instrumen penilaian dapat berupa tes tertulis, tes lisan, lembar pengamatan, pedoman wawancara, tugas rumah, dan sebagainya. Penilaian juga diartikan sebagai kegiatan menafsirkan data hasil pengukuran atau kegiatan untuk memperoleh informasi tentang pencapaian kemajuan belajar peserta didik. Sementara Johnson (2002) menggunakan istilah asesmen untuk menyebut suatu kegiatan mengumpulkan informasi tentang kualitas dan kuantitas perubahan siswa, kelompok, atau guru. Asesmen diartikan sebagai proses untuk menemukan sejumlah deskripsi tingkat karakteristik tertentu yang dimiliki individu.

Evaluasi adalah penilaian yang sistematis tentang manfaat atau kegunaan suatu objek (Mehrens & Lehmann, 1991). Dalam melakukan evaluasi terdapat *judgement* untuk menentukan nilai suatu program yang sedikit banyak mengandung unsur subjektif. Evaluasi memerlukan data hasil pengukuran dan informasi hasil penilaian yang memiliki banyak dimensi, seperti kemampuan, kreativitas, sikap, minat, keterampilan, dan sebagainya. Oleh karena itu, dalam kegiatan evaluasi, alat ukur yang digunakan juga bervariasi bergantung pada jenis data yang ingin diperoleh.

Fungsi evaluasi dalam pembelajaran sangat penting sebagai dasar pengambilan keputusan. Tiga hal yang sangat penting dalam evaluasi meliputi: informasi, interpretasi, dan pengambilan keputusan (Richards, 1996:4). Informasi mengenai proses belajar mengajar biasanya kurang memiliki makna. Untuk dasar pengambilan keputusan, informasi tersebut harus diinterpretasi sehingga memiliki makna.

Uraian di atas menyimpulkan bahwa pengukuran, penilaian, dan evaluasi bersifat bertahap (hierarkis); kegiatan dilakukan secara berurutan, dimulai dengan pengukuran, kemudian penilaian, dan terakhir evaluasi. Pengukuran mengacu pada pengamatan yang diekspresikan secara kuantitatif dan biasanya pengukuran dapat menjawab pertanyaan “seberapa banyak”. Pengukuran merupakan kegiatan awal dalam proses penilaian. Pengukuran adalah suatu istilah umum yang mengacu pada penentuan hasil atau karakteristik sesuatu dengan alat tertentu.

Asesmen dan pengukuran merupakan salah satu prosedur yang dapat digunakan untuk melakukan penilaian. Yang dimaksud prosedur disini adalah langkah atau cara yang dipakai dalam penilaian terutama yang menyangkut kegiatan mengumpulkan data atau informasi yang diperlukan untuk mengambil keputusan dalam penilaian. Salah satu bagian dari proses melakukan penilaian adalah mengumpulkan data. Data yang diperoleh dapat berupa angka-angka atau pernyataan-pernyataan (deskriptif). Jika data yang diperlukan berupa angka-angka prosedur yang digunakan adalah pengukuran. Asesmen lebih mencakup karakteristik kualitatif maupun kuantitatif. Cakupan asesmen lebih luas daripada pengukuran.

Pengukuran lebih berorientasi pada deskripsi kuantitatif, sementara asesmen berorientasi pada deskripsi kuantitatif dan kualitatif.

Isitilah lain yang dikaitkan dengan tes, penilaian, pengukuran, dan evaluasi ialah pengujian (examination). Menurut Hornby (1995:398) pengujian berarti tes yang dilaksanakan secara formal untuk mengetahui sesuatu atau pengetahuan tentang suatu pengetahuan tertentu, misalnya menggunakan pertanyaan tertulis atau melalui praktik. Tujuan pengujian ialah untuk mengetahui kecakapan, mutu, atau kompetensi tertentu dari peserta ujian. Dalam praktik, pelaksanaan tes, penilaian, pengukuran, dan evaluasi diadakan melalui pengujian. Selama dan setelah pembelajaran berlangsung, guru memerlukan umpan balik untuk mengetahui sejauh mana pembelajar telah atau belum menguasai materi ajar. Jadi tes merupakan alatnya, penilaian dan pengukuran ialah proses pengumpulan datanya, pengujian merupakan proses pelaksanaan tesnya, dan evaluasi merupakan proses pengambilan simpulan dan putusan mengenai hasil pengujian.

### **Penilaian Informal dan Penilaian Formal**

Salah satu cara untuk membedakan makna tes, penilaian, dan pengajaran ialah dengan membedakan penilaian informal dan penilaian formal. Penilaian informal bisa dibuat dalam beberapa bentuk, mulai dari kebetulan, komentar dan jawaban tidak terencana, sampai pada membimbing dan balikan langsung dari siswa. Misalnya, guru mengatakan: "*Pekerjaan yang bagus! Sempurna*".

Penilaian informal tidak berhenti di sana. Penilaian informal yang bagus terintegrasi di dalam tugas-tugas kelas yang dirancang untuk menunjukkan kinerja tetapi sengaja tidak diberi skor dan tidak digunakan untuk menentukan tingkat kompetensi siswa. Contoh penilaian jenis ini ialah komentar dalam margin paper untuk merespon draft yang dibuat siswa, saran untuk mengucapkan suatu kata secara lebih baik, strategi mengatasi kesulitan membaca, dan menunjukkan bagaimana membuat catatan kuliah agar bisa diingat lebih baik.

Penilaian formal ialah latihan atau prosedur yang secara spesifik dirancang untuk menunjukkan kemampuan dan pengetahuan. Penilaian ini bersifat sistematis, disusun menggunakan teknik sampling yang terencana untuk memberikan penghargaan terhadap apa yang sudah dicapai siswa. Apakah penilaian formal sama dengan tes? Semua tes ialah penilaian formal, tetapi tidak setiap penilaian formal bisa disebut tes. Misalnya, menggunakan portfolio atau jurnal hasil belajar siswa untuk penilaian bisa disebut penilaian formal, tetapi penggunaan portfolio atau jurnal tersebut tidak bisa disebut tes. Mengamati frekuensi partisipasi siswa dalam kelas bisa disebut melakukan penilaian formal, tetapi pengamatan tersebut juga tidak bisa disebut tes. Tes selalu dikerjakan dalam rentang waktu tertentu dan menggambarkan sampel dari sejumlah perilaku.

### **Penilaian Formatif dan Sumatif**

Teknik lain untuk membedakan tes dan penilaian ialah dilihat berdasarkan fungsi penilaian: bagaimana prosedur tersebut digunakan? Dua fungsi yang biasanya digunakan ialah penilaian formatif dan sumatif. Umumnya penilaian dalam kelas ialah sumatif: mengevaluasi siswa ketika mereka dalam proses membentuk kompetensi dan skill selama dalam pembelajaran. Kunci pembentukan di sini ialah penyampaian oleh guru dan internalisasi oleh siswa terhadap balikan yang sesuai. Untuk semua tujuan praktis, semua jenis penilaian informal ialah formatif. Penilaian ini terjadi sebagai fokus utama perkembangan penguasaan bahasa pembelajar. Ketika guru memberi komentar atau saran terhadap kesalahan siswa, balikan tersebut dimaksudkan untuk meningkatkan kemampuan berbahasa pembelajar.

Penilaian sumatif bertujuan untuk mengukur atau meringkas apa yang telah dicapai siswa dan secara khusus dilaksanakan pada akhir pembelajaran atau akhir unit pembelajaran. Penggambaran apa yang sudah diperoleh siswa berimplikasi pada tuntutan melihat ke belakang sejauh mana siswa telah mencapai tujuan tetapi tidak harus menunjukkan bagaimana perkembangan ke depannya. Ujian akhir suatu mata kuliah atau ujian profesiensi umum adalah contoh penilaian sumatif. Satu persoalan

yang merupakan sikap umum terhadap tes ialah pandangan bahwa semua tes (quiz, review periodik, tes tengah semester) dianggap sumatif.

### **Tes Kriteria Norma dan Kriteria Patokan**

Satu lagi dikotomi yang penting untuk dijelaskan dalam terminologi penilaian ialah perbedaan antara tes berdasar kriteria norma dan tes berdasar kriteria patokan (kriteria). Dalam tes berdasar kriteria norma, skor setiap peserta tes diinterpretasi berdasarkan nilai rata-rata, median, standar deviasi, atau persentil. Tujuan tes jenis ini ialah menempatkan peserta tes ke dalam rentangan matematis dalam bentuk susunan ranking. Skor selalu dilaporkan dalam bentuk angka dan persentil, misalnya persentase. Contoh tes berdasar kriteria norma ialah tes terstandar seperti *Scholastic Aptitude Test* (SAT) atau TOEFL. Tes jenis ini harus baku, jawaban ditentukan lebih dulu dalam format yang bisa diskor secara cepat dengan biaya minimum. Dalam tes jenis ini, penggunaan biaya yang hemat dan efisiensi menjadi perhatian penting.

Tes kriteria patokan dirancang untuk memberikan balikan pada peserta tes berupa peringkat dalam bentuk huruf dalam mata kuliah tertentu atau tujuan pembelajaran tertentu. Tes kelas yang melibatkan peserta hanya dalam satu kelas dan dihubungkan dengan kurikulum ialah contoh tes berdasar kriteria patokan. Di sini, kadang guru memerlukan bantuan administrator tes untuk menentukan balikan yang bermanfaat dan sesuai dengan siswa yang oleh Oller (1979:52) disebut nilai pembelajaran. Dalam tes berbasis kriteria, distribusi skor siswa yang disajikan dalam bentuk kontinum kurang menjadi perhatian asalkan tes tersebut mengukur tujuan yang sesuai. Dalam penilaian bahasa yang pesertanya terdiri dari guru kelas dan guru yang sedang menjalani pelatihan, dan penekanan tesnya pada penilaian berbasis kelas, penggunaan tes kriteria patokan lebih diperlukan dibanding tes kriteria norma.

### **PENDEKATAN DALAM TES BAHASA: SEJARAH SINGKAT**

Sejarahnya, trend dan penggunaan tes bahasa mengikuti perubahan metodologi mengajar. Misalnya pada 1950an ketika era behaviorisme dan perhatian

pembelajaran diarahkan pada analisis kontrastif, tes difokuskan pada elemen khusus suatu bahasa seperti fonologi, grammar, leksikon, yang dikontraskan dengan bahasa lain. Pada era 1970an dan 1980an, teori komunikatif bidang bahasa membawa pengaruh pada penggunaan bahasa secara terintegrasi yang menganggap penggunaan bahasa yang komunikatif secara menyeluruh dipandang lebih penting dibanding dengan penggunaan elemen linguistik. Dewasa ini, pengembang tes masih tertantang untuk menemukan tes yang lebih autentik, valid yang mendorong interaksi senyatanya.

### **Tes Diskrit dan Integratif**

Pendekatan tes bahasa mengacu pada dua jenis yang diperdebatkan pada 1970an dan awal 1980an. Pendekatan tersebut sampai sekarang masih umum digunakan yang disebut dengan tes diskrit dan tes integratif (Oller, 1979). Tes diskrit disusun berdasarkan asumsi bahwa bahasa bisa dipecah-pecah ke dalam bagian-bagian lebih kecil dan setiap bagian bisa dites secara terpisah. Komponen tersebut antara lain: menyimak, berbicara, membaca, dan menulis dan berbagai unit bahasa yang lebih kecil lagi seperti fonologi, grafologi, morfologi, leksikal, sintaksis, dan wacana. Semua jenis tes profisiensi tersebut harus bisa merepresentasikan keempat skill bahasa dan komponen linguistik lain sebanyak mungkin.

Pendekatan tersebut menyebabkan dekontekstualisasi yang sering menyebabkan peserta tes bingung. Ketika era lain datang yang menekankan pada komunikasi, autentisitas, dan konteks, pendekatan baru mulai dicari. Menurut Oller (1979) kompetensi bahasa merupakan gabungan kemampuan interaksi yang tidak bisa dites secara terpisah. Klaim ini menegaskan bahwa kompetensi komunikatif bersifat global dan memerlukan integrasi (yang kemudian disebut tes integratif) yang tidak bisa dilaksanakan dalam bentuk tes grammar, membaca, kosa kata dan bentuk-bentuk tes bahasa diskrit yang lain. Sarjana lain (Cziko, 1982 dan Savignon, 1982) mendukung penggunaan tes integratif ini.

Seperti apa sebenarnya tes integratif itu? Dalam sejarahnya, dua jenis tes yang dianggap sebagai tes integratif ialah *cloze test* dan dikte. *Cloze test* berupa teks bacaan terdiri antara 150-300 kata yang pada setiap kata keenam atau kata ketujuh dihapus; peserta tes diminta memberikan kata yang dianggap cocok untuk diisikan pada kata yang dihilangkan tersebut. Oller (1979) mengklaim bahwa hasil *cloze test* merupakan pengukuran yang bagus untuk melihat semua profisiensi. Menurut konstruk teori yang mendasari klaim ini, kemampuan memberikan jawaban yang sesuai untuk mengisi kata yang dihilangkan dalam teks, memerlukan sejumlah kemampuan yang merupakan kemampuan dasar berbahasa: pengetahuan kosa kata, struktur grammar, struktur wacana, keterampilan dan strategi membaca, dan internalisasi penguasaan grammar (mampu memprediksi satu kata yang diperlukan sesuai dengan sekuensinya). Dikatakan, keberhasilan melengkapi item dalam *cloze test* memerlukan semua keterampilan berbahasa yang hakikatnya merupakan kemampuan global berbahasa.

Dikte merupakan sebuah teknik pengajaran bahasa yang sudah dikenal yang perlahan-lahan berubah menjadi teknik tes. Hakikatnya, peserta tes mendengarkan penguji membaca nyaring atau rekaman teks berisi 100-150 kata dan peserta tes diminta menuliskan apa yang mereka dengar secara benar. Proses menyimak biasanya terdiri dari tiga tahap: membaca nyaring tanpa berhenti, membaca nyaring dengan jeda yang panjang dalam setiap frase (untuk memberi kesempatan peserta tes menuliskan apa yang mereka dengar); dan membaca dengan kecepatan normal untuk memberi kesempatan peserta tes mengecek apa yang mereka tulis.

Para pendukung aliran ini beralasan bahwa dikte merupakan tes integratif sebab tes ini memerlukan kompetensi grammar dan kompetensi wacana untuk digunakan pada kemampuan berbahasa yang lain. Keberhasilan menjawab tes dikte memerlukan menyimak yang cermat, reproduksi tertulis apa yang mereka dengar, daya ingat yang efisien dalam jangka pendek, dan sampai tahap tertentu penguasaan aturan kebahasaan untuk membantu daya ingat jangka pendek. Selanjutnya, hasil tes dikte cenderung menunjukkan korelasi yang kuat dengan tes profisiensi lain. Tes

dikte selalu merupakan tes berbasis kelas karena proses administrasi tes dikte dalam jumlah besar sangatlah tidak praktis dilihat dari teknik skoringnya. Reliabilitas kriteria skoring tes dikte bisa ditingkatkan melalui pilihan ganda atau skoring berdasarkan satu jawaban benar.

Pengguna metode tes integratif selanjutnya menggunakan apa yang disebut dengan *hipotesis kesatuan perilaku (unitary trait hypothesis, UTH)* yang menyarankan adanya kemampuan bahasa yang tidak bisa dipisah-pisahkan: bahwa kosa kata, grammar, fonologi, empat skill, dan bentuk diskrit lain dalam bahasa tidak bisa dipisahkan dalam penggunaan kemampuan berbahasa. UTH berkeyakinan bahwa ada faktor umum kemampuan berbahasa yang menunjukkan bahwa faktor diskrit merupakan bagian dari kesatuan umum tersebut.

Pihak lain membantah keras pandangan UTH. Dalam penelitian yang dilakukan di Brazil dan Filipina, Farhady (1982) menemukan perbedaan kinerja yang signifikan dan sangat bervariasi dalam tes profisiensi ESL, tergantung pada Negara asal peserta tes, bidang ilmu yang dipelajari, dan lulusan sarjana atau bukan. Misalnya, peserta dari Brazil memperoleh skor sangat rendah dalam tes menyimak dan memperoleh skor relatif tinggi dalam tes membaca. Peserta dari Filipina yang skor dalam 5 atau 6 komponen tes dianggap lebih tinggi dibanding Brazil, sebenarnya memiliki skor membaca lebih rendah dibanding Brazil. Akhirnya, setelah melihat bukti tersebut, Oller (1983) mengoreksi pendapatnya dan mengatakan bahwa UTH terbukti keliru.

### **Tes Bahasa Komunikatif**

Pertengahan 1980an bidang tes bahasa dihadapkan pada UTH dan mulai mengkaji bentuk-bentuk tes komunikatif. Bachman dan Palmer (1996:9) mengemukakan prinsip fundamental dalam tes bahasa dengan melihat hubungan antara tes kinerja berbahasa dengan penggunaan bahasa. Agar tes unsur khusus bahasa bermanfaat untuk tujuan yang dikehendaki, tes kinerja harus menunjukkan kemampuan yang bisa didemonstrasikan dalam bentuk penggunaan

bahasa dalam situasi non-tes. Permasalahan yang dihadapi ahli tes bahasa ialah tugas dalam tes bersifat tiruan, dipaksakan, dan jarang digunakan dalam kehidupan senyatanya. Seperti dikatakan Weir (1990:6) tes integratif misalnya *cloze test* hanya menjelaskan kemampuan linguistik peserta tes. Tes integratif tidak menjelaskan apa pun yang secara langsung menunjukkan kemampuan kinerja peserta tes.

Dan bersamaan dengan temuan tentang autentisitas diluncurkan, perancang tes memfokuskan pada kinerja komunikatif. Mengikuti pandangan Canale dan Swain (1980) tentang Kompetensi Komunikatif, Bachman (1990) mengusulkan model kompetensi bahasa yang terdiri dari kompetensi organisasi dan pragmatik, yang berturut-turut dikelompokkan ke dalam komponen grammar dan teks, dan ke dalam komponen ilokusi dan sosiolinguistik. Bachman dan Palmer (1996:70f) juga menekankan pentingnya kompetensi strategik (kemampuan menggunakan strategi komunikasi untuk mengatur cara menyampaikan ujaran untuk menguatkan pengaruh retorika ujaran) dalam proses komunikasi. Semua unsur dalam model tersebut khususnya kemampuan pragmatik dan strategik, perlu dimasukkan ke dalam konstruk tes bahasa dan ke dalam kinerja senyatanya yang diperlukan oleh peserta tes.

Tes komunikatif memunculkan tantangan bagi perancang tes. Penyusun tes mulai menyusun jenis-jenis tugas yang senyatanya yang benar-benar diperlukan peserta tes untuk ditampilkan. Jelas bahwa konteks untuk menerapkan tugas (task) dalam tes, sangat luas cakupannya dan sampling dalam tes harus divalidasi oleh pengguna bahasa dalam situasi senyatanya. Weir (1990:11) mengingatkan bahwa untuk mengukur profisiensi berbahasa harus memperhatikan aspek: di mana, kapan, bagaimana, dengan siapa, dan mengapa bahasa digunakan, menggunakan topik apa, dan apa pengaruhnya. Akhirnya bidang penilaian menjadi lebih berkaitan dengan autentisitas tugas dan keaslian teks.

### **Penilaian Berbasis Kinerja**

Dalam program pembelajaran bahasa di seluruh dunia, perancang tes dewasa ini memfokuskan diri pada agenda tes berpusat pada siswa (Alderson, 2001, 2002).

Selain hanya menggunakan tes tertulis dengan jawaban selektif, tes bahasa berbasis kinerja secara khusus melibatkan tes lisan, tes tertulis, tes uraian, kinerja terintegrasi, kinerja kelompok, dan jenis tugas terintegrasi lainnya. Dipastikan, tes jenis ini sangat menyita waktu sehingga biayanya mahal, tetapi upaya tersebut dianggap lebih langsung karena siswa dinilai ketika mereka menampilkan tuntutan tugas yang senyatanya. Dalam istilah teknis, validitas isi yang lebih tinggi bisa dicapai karena pembelajar dinilai ketika mereka sedang dalam proses menampilkan perilaku linguistik yang dikehendaki.

Dalam konteks tes bahasa Inggris, tes berbasis kinerja menimbulkan kesulitan untuk membedakan antara penilaian informal dan penilaian formal. Dengan sedikit mengurangi perhatian terhadap tes yang disusun secara formal, dan perhatian lebih sedikit terhadap evaluasi ketika siswa sedang menampilkan berbagai tugas, kita selangkah lebih maju untuk menemukan tujuan tes berbasis kinerja.

Salah satu karakteristik dari banyak tes berbasis kinerja ialah keberadaan tugas yang terintegrasi. Dalam konteks ini, penilaian melibatkan penampilan pembelajar secara nyata perilaku yang akan dinilai. Dalam tugas terintegrasi, peserta tes dinilai dalam penampilannya berbicara, bertanya, menjawab atau gabungan ketika menyimak dan berbicara, dan dalam mengintegrasikan membaca dan menulis. *Paper-and-pencil test* jelas tidak menunjukkan kinerja komunikatif tersebut.

Contoh penting prosedur penilaian bahasa interaktif ialah wawancara lisan. Peserta tes diminta mendengarkan secara cermat orang lain dan diminta menjawab yang sesuai. Jika perhatian diarahkan pada proses penyusunan tes, bahasa yang ditampilkan yang ditunjukkan secara suka rela oleh siswa secara individu dan bermakna, hal-hal itu mendekati kenyataan otentik penggunaan bahasa yang senyatanya.

## **ISU-ISU MUTAKHIR DALAM TES BAHASA**

Rancangan rubrik penilaian komunikatif dan penilaian berbasis kinerja menunculkan tantangan bagi ahli tes dan para guru. Upaya untuk memperbaiki

penggunaan tes demikian diikuti dengan isu yang mendorong pemahaman penggunaan penilaian yang efektif. Tiga isu yang kini berkembang ialah: pengaruh teori baru bidang intelegensi terhadap tes bidang industri, kedatangan apa yang disebut dengan penilaian alternatif, dan peningkatan popularitas penggunaan tes berbasis komputer.

### **Pandangan Baru dalam Intelegensi**

Intelegensi secara ketat dianggap sebagai kemampuan untuk menampilkan pemecahan masalah di bidang bahasa dan berpikir logis secara matematis. Konsep IQ (intelligence quotient) telah lazim digunakan di Barat dan mempengaruhi penggunaan tes hampir pada semua abad. Karena kecerdasan secara umum diukur menggunakan waktu, tes diskrit yang berisi hirarki item terpisah, mengapa setiap bidang kajian tidak diukur dengan cara yang sama? Dalam beberapa waktu, kita hidup dalam dunia yang distandarkan, tes berbasis kriteria norma yang disajikan dalam format tes pilihan ganda yang berisi logika majemuk yang dibatasi dalam item dan banyak di antaranya tidak autentik.

Penelitian bidang intelegensi yang dilakukan oleh psikolog seperti Howard Gardner, Robert Sternberg, dan Daniel Goleman telah menurunkan penggunaan konsep spikometrik. Gardner (1983, 1999) misalnya mengembangkan pandangan tradisional tentang intelegensi ke dalam tujuh komponen. Dia menerima konsep lama intelegensi kecerdasan linguistik dan kecerdasan matematis yang selama ini digunakan sebagai tes terstandar IQ, tetapi dia memasukkan lima kerangka pemikiran lainnya dalam teori yang disebut kecerdasan majemuk: (1) kecerdasan spasial (kemampuan menemukan jalan keluar dari lingkungan sekitar, kemampuan membentuk keasn mental dari suatu realitas), (2) kecerdasan musical (kemampuan mempersepsi dan menciptakan irama dan bentuk-bentuk ritme), (3) kecerdasan gerak tubuh-kinestesik (menampilkan gerak yang indah, atletis), (4) kecerdasan interpersonal (kemampuan memahami pihak lain dan bagaimana mereka merasakan

dan berinteraksi secara efektif dengan mereka), dan (5) kecerdasan intrapersonal (kemampuan memahami diri sendiri untuk membentuk pemahaman identitas diri).

Robert Sternberg (1988, 1997) juga menemukan teritorial baru di bidang intelegensi dalam bentuk berpikir kreatif dan strategi manipulatif sebagai bagian dari intelegensi. Semua orang cerdas tidak selalu bertindak secara cepat, berpikir reaktif. Mereka bahkan sangat inovatif dan bisa berpikir melebihi batas-batas normal yang dirancang dalam tes, tetapi mereka memerlukan waktu yang cukup untuk memproses kreativitas tersebut. Bentuk kecerdasan lain ditemui dalam diri mereka yang mengetahui bagaimana memanipulasi lingkungan, yaitu orang lain. Ahli debat, politisi, sales yang sukses, pembicara ahli, dan artis semuanya adalah orang cerdas dalam memanipulasi kemampuan meyakinkan orang lain untuk mengikuti pendapatnya, memilih mereka, membeli dagangannya, atau melakukan sesuatu yang bahkan mereka tidak berniat melakukan.

Lebih mutakhir lagi, Daniel Goleman (1995) yang memperkenalkan konsep EQ (emotional quotient, kecerdasan emosional) telah mencengangkan kita untuk mengakui bahwa pentingnya faktor emosi dalam proses kognitif. Mereka yang mampu mengendalikan emosinya –terutama emosi yang membahayakan- cenderung lebih mampu memproses intelegensinya secara penuh. Kamarahan, kekecewaan mendalam, terhina, kurang percaya diri, dan perasaan-perasaan lain dengan mudah bisa merusak kinerja yang bagus dalam mengerjakan tugas sehari-hari termasuk pola pemecahan masalah yang memerlukan tingkat berpikir lebih tinggi.

Konsep baru tentang intelegensi ini belum bisa diterima sepenuhnya oleh kalangan akademisi (misalnya White, 1998). Namun, daya tarik intuitifnya telah mewarnai dekade 1990an yang menentang kebebasan dan tanggungjawab dalam agenda testing. Sejalan dengan reformasi pendidikan pada masa itu (Amstrong, 1994) konsep-konsep itu membantu membebaskan dari ketergantungan pada tes-tes yang terikat waktu, diskrit dan analitik dalam mengukur kemampuan berbahasa. Kita didorong untuk terus berjuang melawan tirani objektivitas dan pendekatan impersonal yang menyertainya. Tetapi kita juga berharap pada

tanggungjawab menyeluruh penggunaan skill berbahasa, proses belajar, dan kemampuan menegosiasikan makna. Tantangan kita ialah mengetes interpersonal, kreativitas, kemampuan interaktif, dan sejenisnya untuk memperoleh kepercayaan terhadap subjektivitas dan institusi.

### **Penilaian Tradisional dan Alternatif**

Secara implisit dalam deskripsi penilaian kelas berbasis kinerja terdapat kecenderungan penggunaan design tes tradisional dan alternatifnya yang dianggap lebih otentik dalam memunculkan komunikasi yang bermakna. Dua hal perlu ditekankan di sini. Pertama, konsep penilaian tradisional dan penilaian alternatif nampak berlebihan sehingga harus disikapi dengan hati-hati. Banyak bentuk penilaian berada di antara keduanya, dan beberapa yang lainnya menggabungkan keduanya. Kedua, jelas bahwa konsep itu menimbulkan bias dalam memandang penilaian alternatif, dan orang tidak boleh salah mengartikan. Brown dan Hudson (1998) menegaskan bahwa tradisi penilaian yang ada harus dihargai dan digunakan sesuai fungsinya. Perlu juga ditekankan di sini bahwa waktu yang lama dan anggaran yang lebih banyak harus disiapkan untuk tes yang dilaksanakan dalam bentuk subjektif, lebih bersifat individu, dan lebih interaktif guna mendapatkan balikan.

### **Penilaian Berbasis Komputer**

Tahun-tahun terakhir ini banyak digunakan penilaian yang memungkinkan peserta tes menjawab menggunakan komputer. Beberapa tes berbasis komputer merupakan tes dalam skala kecil yang tumbuh di rumah yang tersedia di website. Yang lainnya berbentuk terstandar, berskala besar yang melibatkan ribuan bahkan puluhan ribu peserta tes. Siswa menerima perintah atau petunjuk lisan atau tertulis dalam komputer dan diminta mengetik jawabannya. Hampir semua jenis tes berbasis komputer bersifat baku, menggunakan jawaban pendek; namun tes seperti TOEFL menawarkan tes esai tertulis yang penilaiannya dilakukan oleh manusia.

Jenis tes berbasis komputer yang tersedia saat ini ialah tes komputer adaptif (CAT). Dalam CAT, siswa menerima sejumlah pertanyaan yang sesuai dengan spesifikasi tes yang umumnya sesuai dengan tingkat kemampuan peserta tes. CAT dimulai dengan pertanyaan dengan tingkat kesulitan sedang. Peserta menjawab pertanyaan tersebut dan komputer menilainya, dan berdasarkan jawaban dan nilai yang diperoleh tersebut, komputer memproses jenis pertanyaan mana yang akan ditampilkan kemudian. Jika jawaban benar, komputer akan memunculkan pertanyaan yang kurang lebih sama atau lebih sulit tingkat kesulitannya. Jika jawaban salah, komputer akan memunculkan pertanyaan yang tingkat kesulitannya lebih rendah. Dalam CAT, peserta tes hanya melihat satu pertanyaan saja dan komputer akan mengoreksi jawabannya sebelum pertanyaan lain dimunculkan. Dengan demikian, peserta tes tidak bisa menunda menjawab soal dan begitu peserta tes sudah dikonfirmasi jawabannya, dia tidak bisa melihat lagi pertanyaannya atau bagian tes sebelumnya.

Tes berbasis komputer memiliki keunggulan (1) bersifat penilaian kelas, (2) tes yang bersifat perintah mandiri untuk berbagai aspek bahasa, (3) praktis untuk pelaksanaan tes terstandar, (4) beberapa bersifat individu, misalnya CAT, (5) bisa digunakan untuk tes terstandar dalam skala besar. Kelemahan tes jenis ini ialah (1) kurang aman dan terbuka peluang untuk nyontek, (2) secara sengaja merupakan quiz yang tumbuh di rumah yang sulit divalidasi, (3) lebih banyak menggunakan format pilihan ganda, (4) pertanyaan uraian kurang disukai karena memerlukan pengoreksi manusia, (5) tidak ada interaksi manusia.

Penilaian bersifat integral dengan siklus belajar-mengajar. Dalam kurikulum dan komunikatif, penilaian bersifat tetap. Tes sebagai bagian dari penilaian bisa memberikan otentisitas, motivasi, dan balikan pada pembelajar. Tes merupakan komponen penting dalam pelaksanaan kurikulum yang berhasil dan menjadi salah satu bagian dari proses belajar. Penilaian memiliki beberapa prinsip: (1) penilaian periodik, berbentuk formal dan informal berfungsi meningkatkan motivasi dengan menunjukkan kemajuan siswa; (2) penilaian yang sesuai membantu penguatan dan

daya tahan informasi; (3) penilaian bisa menegaskan kekuatan dan bagian-bagian penting yang memerlukan pekerjaan lanjutan; (4) penilaian bisa memberikan tambahan modul secara periodik untuk melengkapi kurikulum; (5) penilaian bisa mendorong kemandirian siswa melalui penilaian mandiri; (6) penilaian bisa mendorong siswa menentukan tujuan belajarnya sendiri; (7) penilaian bisa membantu evaluasi efektivitas pembelajaran.

### **PERMASALAHAN**

Uraian di atas menunjukkan terdapat perkembangan baru yang menarik di bidang penilaian bahasa. Perkembangan baru tersebut ada yang bersifat mengoreksi, memodifikasi, menggabungkan, dan memunculkan hal baru. Sejalan dengan hal tersebut, penerjemah mengajukan tiga permasalahan berikut:

- 1) Apakah perkembangan baru di bidang penilaian bahasa sebagaimana dikemukakan dalam uraian bab ini sudah sesuai dengan model penilaian di sekolah?
- 2) Apakah konsep-konsep penilaian dalam uraian bab ini sudah sesuai dengan perkembangan sistem penilaian di Indonesia?
- 3) Apakah para guru sudah menguasai model penilaian sebagaimana dikemukakan dalam uraian bab ini untuk diterapkan dalam penilaian berbasis kinerja dan penialain berbasis komputer?

### **PEMBAHASAN**

Bagian pembahasan ini dimaksudkan untuk menjawab empat permasalahan di atas. Jawaban permasalahan sengaja tidak dicantumkan dalam simpulan karena komentar dan uraian luas untuk mengkaji permasalahan menurut hemat penerjemah lebih sesuai ditempatkan di bagian pembahasan, bukan pada bagian simpulan sebagaimana kerangka tulisan yang ditawarkan dalam tugas ini.

### **Model Penilaian Bahasa di Sekolah**

Uraian dalam bab ini bisa dianggap segar dan kreatif. Brown (2004) membuat definisi yang cerdas mengenai tes, penilaian, dan pengajaran. Uraian ini menegaskan bahwa yang terpenting dalam proses pembelajaran ialah penilaian. Penilaian dilakukan selama proses pembelajaran berlangsung baik secara formal atau informal. Dalam melaksanakan penilaian (assessment) guru harus merancang secara cermat tetapi menerapkannya dengan variasi yang penuh pertimbangan. Ini memberi motivasi pada guru agar selalu cermat dalam melaksanakan pembelajaran. Penilaian tidak selalu diartikan memberikan nilai untuk dasar keputusan tentang status siswa tetapi juga memberi motivasi dan mendorong proses belajar mencapai tujuan pembelajaran yang ditargetkan. Tes, menurut Brown (2004) ialah metode mengambil nilai. Tes memiliki ciri dilaksanakan dalam waktu khusus, menguji pengetahuan yang sudah diajarkan, dan berisi sampel perilaku belajar dalam ranah tertentu. Karena itu, tes tidak sama dengan penilaian baik secara substansi maupun dalam cara melaksanakannya. Dalam keseluruhan pengajaran, guru selalu mengadakan penilaian, tetapi hanya mengadakan tes dalam waktu yang sudah ditentukan saja.

Definisi ini memberi penegasan pada jenis tes yang dalam kurun waktu sebelum KTSP diluncurkan pada 2006 menjadi begitu populer. Saat itu, tes yang banyak digunakan ialah tes tertulis berbentuk objektif umumnya pilihan ganda. Tes esai uraian kurang diminati apalagi tes berbentuk karya tulis misalnya paper. Pelaksanaan tes juga lebih mengacu pada jenis tes formatif untuk mengukur pencapaian satu atau beberapa unit bahan yang sudah diajarkan dan tes sumatif yang dilaksanakan pada akhir semester. Dengan demikian, penilaian (assessment) kurang dikenal.

Jika keadaan ini dihubungkan dengan kondisi pembelajaran di Indonesia, bisa disimpulkan bahwa guru sudah kenal jenis-jenis penilaian tetapi belum tahu bagaimana menyusun dan menggunakannya sesuai dengan kebutuhan pembelajaran di kelas. Penerjemah menemukan tiga fakta untuk menjawab permasalahan tersebut.

Pertama, guru lebih mengenal penilaian dalam bentuk format tes berupa tes pilihan ganda. Hal ini disebabkan tradisi penilaian dalam sistem persekolahan di Indonesia masih mengacu pada penilaian menggunakan pendekatan tes. Tes berbentuk pilihan ganda dianggap lebih mudah dan lebih praktis. Selain itu, hanya sedikit jumlah guru yang menguasai teknik penyusunan tes.

Kedua, penilaian dalam konteks *assessment* belum banyak dikenal sehingga tidak dilaksanakan karena belum ada model yang bisa ditiru. Hal ini terkait dengan metode mengajar yang digunakan di kelas. Walaupun model pembelajaran inovatif seperti *direct approach*, *self-motivated learning*, *problem based learning*, dengan variasi *cooperative learning* yang mengedepankan Pembelajaran Aktif, Kreatif, Inovatif dan Menyenangkan (PAKEM/PAIKEM), praktiknya di dalam kelas guru tetap menggunakan model pembelajaran tradisional. Penilaian dalam bidang bahasa khususnya bahasa Inggris tetaplah menguji grammar, menjawab pertanyaan dalam reading teks, atau menampilkan skenario *conversation* yang secara artificial ditentukan oleh guru. Penilaian dalam konteks *assessment* atau *classroom based assessment* atau *performance based assessment* belum dilakukan.

Ketiga, kebijakan pemerintah dalam menerapkan KTSP yang di dalamnya menekankan penilaian berbasis kelas dengan *authentic assessment* belum dikuasai guru. Berbagai pelatihan kurikulum, menyusun bahan ajar, penilaian dan strategi mengajar sudah banyak dilakukan sejak 2006. Namun pelatihan itu lebih berorientasi pada pengisian format. Di sekolah-sekolah dari tingkat SD, SMP dan SMA, penilaian portofolio dan penilaian berbasis kelas juga sudah dilaksanakan oleh setiap guru. Penilaian untuk menguji satu Kompetensi Dasar (KD) atau beberapa KD yang disebut ujian blok juga sudah diwajibkan, tetapi penekanan yang dikerjakan guru ialah mengisi format penilaian untuk rapor. Sedikit sekali pembekalan yang menekankan guru mandiri dan bisa merancang, membuat, melaksanakan, mengoreksi, dan menafsirkan penilaian. Guru banyak dibatasi oleh penilaian yang terikat pada penyelesaian mengejar nilai Ujian Nasional (UN).

Dengan demikian bisa dikatakan bahwa macam-macam penilaian, model penilaian, dan bentuk-bentuk pelaksanaan penilaian sudah dikenal di Indonesia dan sudah dituangkan dalam dokumen nasional pendidikan. Bahkan aturan mengenai penilaian juga sudah dituangkan dalam UU dan Peraturan Pemerintah. Tetapi dalam taraf implementasi guru hanya mengenal jenis penilaian yang biasa digunakan sehari-hari saja. Guru juga belum dibekali kemampuan mendesign penilaian secara mandiri.

### **Sistem Penilaian di Indonesia**

Pengajaran-tes-penilaian merupakan rangkaian tak terpisahkan. Pengajaran dimulai dengan pembuatan tujuan pembelajaran, perancangan bahan ajar, penerapan bahan ajar, dan evaluasi. Dengan demikian, evaluasi tidak terpisahkan dalam pembelajaran. Pandangan ini memberi kecenderungan bahwa penilaian hanya dilakukan pada akhir pembelajaran.

Konsep ini kurang sejalan dengan sistem penilaian dalam bentuk *assessment*. *Assessment* dilaksanakan selama proses pembelajaran berlangsung dan dapat dilaksanakan secara sengaja dalam bentuk tes atau secara tidak formal dalam bentuk komentar atau petunjuk untuk meningkatkan motivasi siswa sampai siswa memperoleh keterampilan yang diharapkan. Dengan demikian, penilaian sepenuhnya menjadi hak mutlak guru.

Dalam pelaksanaan sehari-hari, sistem penilaian sudah mengarah pada pemberian kewenangan pada guru. Model penilaian berbasis kelas menggunakan portofolio, proyek, kinerja juga sudah dilaksanakan. Dalam menentukan evaluasi nasional, pemerintah juga sudah menyediakan porsi yang memadai: kelulusan ditentukan tidak hanya dari UN tetapi juga Ujian Sekolah (US). Namun penerapan kebijakan ini belum maksimal. Beberapa fakta menunjukkan bahwa sekolah yang tidak berhasil meluluskan siswanya dalam UN sampai 95% kepala sekolahnya bahkan kepala dinasya dicopot. Akibatnya, sekolah lebih mengutamakan UN sedangkan US digunakan sebagai syarat saja. Kondisi demikian semakin

memprihatinkan untuk sekolah-sekolah di pedesaan atau pedalaman yang jauh dari fasilitas dan lingkungan belajar yang standar.

Jawaban ini menegaskan bahwa sistem penilaian di Indonesia sudah mengikuti perkembangan penilaian di dunia pendidikan umumnya, namun pelaksanaannya terkendala oleh faktor non-penilaian yang ada di luar teknik penilaian.

Penilaian dalam KTSP adalah penilaian berbasis kompetensi, yaitu bagian dari kegiatan pembelajaran yang dilakukan untuk mengetahui pencapaian kompetensi peserta didik yang meliputi pengetahuan, keterampilan, dan sikap. Penilaian dilakukan selama proses pembelajaran dan/atau pada akhir pembelajaran. Fokus penilaian pendidikan adalah keberhasilan belajar peserta didik dalam mencapai standar kompetensi yang ditentukan. Pada tingkat mata pelajaran, kompetensi yang harus dicapai berupa Standar Kompetensi (SK) mata pelajaran yang selanjutnya dijabarkan dalam Kompetensi Dasar (KD). Untuk tingkat satuan pendidikan, kompetensi yang harus dicapai peserta didik adalah Standar Kompetensi Lulusan (SKL).

Kualitas pendidikan sangat ditentukan oleh kemampuan satuan pendidikan dalam mengelola proses pembelajaran. Penilaian merupakan bagian yang penting dalam pembelajaran. Dengan melakukan penilaian, pendidik sebagai pengelola kegiatan pembelajaran dapat mengetahui kemampuan yang dimiliki peserta didik, ketepatan metode mengajar yang digunakan, dan keberhasilan peserta didik dalam meraih kompetensi yang telah ditetapkan. Berdasarkan hasil penilaian, pendidik dapat mengambil keputusan secara tepat untuk menentukan langkah yang harus dilakukan selanjutnya. Hasil penilaian juga dapat memberikan motivasi kepada peserta didik untuk berprestasi lebih baik.

Penilaian dalam KTSP menggunakan acuan kriteria. Maksudnya, hasil yang dicapai peserta didik dibandingkan dengan kriteria atau standar yang ditetapkan. Apabila peserta didik telah mencapai standar kompetensi yang ditetapkan, ia dinyatakan lulus pada mata pelajaran tertentu. Apabila peserta didik belum mencapai

standar, ia harus mengikuti program remedial/perbaikan sehingga mencapai kompetensi minimal yang ditetapkan.

Penilaian yang dilakukan harus memiliki asas keadilan yang tinggi. Maksudnya, peserta didik diperlakukan sama sehingga tidak merugikan salah satu atau sekelompok peserta didik yang dinilai. Selain itu, penilaian tidak membedakan latar belakang sosial-ekonomi, budaya, bahasa, gender, dan agama. Penilaian juga merupakan bagian dari proses pendidikan yang dapat memacu dan memotivasi peserta didik untuk lebih berprestasi meraih tingkat yang setinggi-tingginya sesuai dengan kemampuannya.

Ditinjau dari sudut profesionalisme tugas kependidikan, kegiatan penilaian merupakan salah satu ciri yang melekat pada pendidik profesional. Seorang pendidik profesional selalu menginginkan umpan balik atas proses pembelajaran yang dilakukannya. Hal tersebut dilakukan karena salah satu indikator keberhasilan pembelajaran ditentukan oleh tingkat keberhasilan yang dicapai peserta didik. Dengan demikian, hasil penilaian dapat dijadikan tolok ukur keberhasilan proses pembelajaran dan umpan balik bagi pendidik untuk meningkatkan kualitas proses pembelajaran yang dilakukan.

### **Penguasaan Guru Terhadap Model Penilaian**

Penguasaan guru terhadap model-model penilaian yang sesuai dengan model pembelajaran inovatif bisa dikatakan masih kurang. Indikasi kurang bisa dilihat dari praktik sehari-hari dalam penilaian di kelas dan untuk mengisi rapor. Selain itu, data dalam sertifikasi guru yang mengikuti portofolio atau PLPG, program pelatihan untuk guru yang tidak lulus portofolio, menunjukkan bahwa sedikit saja guru yang menguasai teknik mengembangkan penilaian.

Pertama, guru sebelumnya terbiasa dengan model tes objektif misalnya pilihan ganda yang digunakan pada hampir semua tujuan penilaian: formatif, sumatif, ujian akhir. Kini guru harus menyesuaikan kemampuannya berbentuk penilaian berbasis kelas. Penilaian berbasis kelas mengedepankan authentic assessment berupa

portofolio, proyek, *performance*. Bentuk ujian yang cocok dengan *authentic assessment* ini ialah paper. Di sinilah permasalahan itu muncul. Guru sendiri terbiasa membuat paper sehingga mereka kurang menguasai bagaimana rubrik penilaian untuk paper harus dibuat.

Kedua, guru dilatih membuat penilaian berbasis kelas tetapi penekanan pelatihannya ialah pada penguasaan format. Guru diberi format-format yang dianggap bagus dan cocok untuk jenis penilaian tertentu. Guru kemudian dilatih untuk mengisi kolom-kolom dalam format tersebut. Tetapi isi format tersebut memerlukan pemahaman yang baik dan benar tentang hakikat penilaian dan guru tidak diberi informasi yang memadai serta bisa dipertanggungjawabkan. Akibatnya, guru cenderung hanya memindahkan isi format tanpa berani membuat modifikasi atau membuat hal baru yang berbeda.

Ketiga, guru tidak sepenuhnya menguasai materi yang diajarkan dan yang harus dibuat dalam bentuk tes. Ini terkait dengan penguasaan *subject matters* suatu bidang studi. Karena tidak menguasai substansinya, guru merasa takut untuk membuat sesuatu yang baru yang mereka tidak kuasai sebelumnya. Dalam perspektif lain, guru bidang studi Bahasa Inggris di sekolah A misalnya, belum tentu mampu menjawab dengan benar soal-soal dalam UN untuk siswanya. Hal ini menunjukkan bahwa penguasaan guru terhadap substansi penilaian belum maksimal sehingga menghambat penguasaan membuat penilaian.

Keempat, penilaian berbasis kinerja dan penilaian berbasis komputer di Indonesia belum banyak dikembangkan. Penilaian berbasis kinerja terkendala oleh penguasaan bahan ajar, pengembangan tes, dan kebijakan pemerintah yang belum secara khusus mengarah ke penggunaan penilaian jenis ini. Penilaian berbasis komputer secara permanen terkendala oleh kesiapan dana, sarana prasarana, dan SDM yang belum siap. Sejauh ini, pemerintah Indonesia masih memiliki kendala permanen dari aspek logistik pendidikan. Karena itu, pengembangan model penilaian ke arah peningkatan kualitas masih menjadi persoalan yang mendasar.

## SIMPULAN

Uraian dalam bab yang dikaji dan pembahasan yang dikaitkan dengan kondisi empiris di lapangan membawa pada simpulan berikut ini.

- 1) Substansi isi bab yang dikemukakan Brown (2004) ini tergolong *fresh* dan kreatif. Isi bab ini bagus untuk memberikan kerangka pikir bagaimana perkembangan penilaian secara umum dan penilaian bidang bahasa. Brown menguraikan posisi penilaian dengan sangat baik melalui definisi tes, *assessment*, dan *teaching*. Dalam uraian lebih lanjut, Brown juga membedakan jenis tes diskrit dan integratif. Uraian ini secara khusus memberikan penjelasan mengenai perkembangan pandangan bahasa sebagai kompetensi diskrit, sosiolinguistik, wacana, dan strategik sekaligus perkembangan di bidang tes bahasa. Dalam konteks modern, perkembangan isu tes bahasa juga dibahas dikaitkan dengan teori kecerdasan majemuk dalam penguasaan bahasa, tes alternatif berupa non-tes, dan tes berbasis komputer. Pandangan baru tersebut sudah mulai dikenal di Indonesia bahkan pada taraf tertentu pemerintah sudah menerapkan di lapangan. Namun kendala muncul karena hanya sedikit guru yang menguasai. Dalam banyak hal, guru tetap menarapkan pola lama yang tidak sesuai dengan konsep perkembangan tes sekarang. Kesanggupan pemerintah untuk secara konsisten menerapkan kebijakan baru juga masih banyak kendalanya.
- 2) Konsep-konsep penilaian yang dikembangkan di Indonesia sudah memiliki kesamaan dengan uraian bab ini. Ini berarti kebijakan pendidikan di Indonesia sudah mengacu pada perkembangan teori belajar terbaru. Diluncurkannya KTSP dengan model penilaian berbasis kelas yang menggunakan *authentic assessment* sudah merujuk pada perkembangan teori baru tersebut. Namun orientasi pendidikan yang belum mengacu pada peningkatan kualitas belum memberi kontribusi yang memadai. Pelatihan yang diberikan terhadap guru-guru untuk menguasai sistem penilaian ini lebih cenderung bersifat penguasaan format dan belum mengarah pada pembekalan kompetensi sampai guru mandiri.

- 3) Guru sebenarnya belum siap meskipun pemerintah sudah menuangkan kebijakan itu dalam bentuk UU dan PP. Dua jenis penilaian yang nampak masih belum bisa diterapkan di Indonesia ialah penilaian berbasis kinerja dan penilaian berbasis komputer. Faktor penyebabnya terutama logistik yang meliputi kesiapan anggaran, konsistensi kebijakan, dan SDM. Dari segi guru, selain kurang menguasai *subject matters* yang akan dibuat soal untuk penilaian, penguasaan teknik pengembangan penilaian juga masih menjadi persoalan mendasar. Tapi bagaimanapun, pikiran Brown dalam buku ini memberi penyegaran dan inspirasi untuk membuka lebih luas cakrawala sistem penilaian dan perjuangan lebih keras lagi dalam meningkatkan penguasaan penilaian. Ketika Indonesia belum menguasai sepenuhnya suatu sistem, sistem terbaru yang sudah lebih maju sudah dilaksanakan orang lain dengan sangat baik.

#### DAFTAR RUJUKAN

- Brennan, R.L (Ed.). 2006. *Educational Measurement*. Westport, CT: An Imprint of Greenwood Publishing Group Inc.
- Brown, H.D. 2004. *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson Education Inc.
- DePorter B, Reardon M, & Singer-Nourie S. 1999. *Quantum Teaching: Mempraktikkan Quantum Learning di Ruang-Ruang Kelas*. Bandung: Kaifa.
- Douglas, D. 2000. *Assessing Language for Specific Purposes*. Cambridge: Cambridge University Press.
- Gay, LR. 1985. *Educational Evaluation and Measurement: Competencies for Analysis and Application*. Columbus: A Bell & Howel Company.
- Gronlund EN. 1985. *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Company.
- Hill, H. W. 2009. *Theories of Learning: Teori-Teori Pembelajaran, Konsepsi, Komparasi dan Signifikansi*. Bandung: Penerbit Nusa Media.
- Johnson, E.B. 2002. *Contextual Teaching & Learning*. Thousand Oaks, California: Corwin Press Inc.
- Mudjiman, Haris. 2008. *Belajar Mandiri (Self-Motivated Learning)*. Surakarta: UNS Press.
- Muijs, D & Reynold D. 2008. *Effective Teaching: Teori dan Aplikasi*. Yogyakarta: Pustaka Pelajar.
- Mulyasa E. 2009. *Kurikulum Yang Disempurnakan: Pengembangan Standar Kompetensi dan Kompetensi Dasar*. Bandung: Penerbit Rosda.

- Panduan Lengkap KTSP*. 2008. Yogyakarta: Pustaka Yustisia.
- Richards, JC (Ed.). 1996. *Classroom-Based Evaluation in Second Language Education*. Cambridge: Cambridge University Press.
- Sarwiji Suwandi. 2008. *Model Asesmen dalam Pembelajaran*. Surakarta: Panitia Sertifikasi Guru Rayon 13.
- Silberman, M. 2002. *Active Learning: 101 Strategi Pembelajaran Aktif*. Yogyakarta: Penerbit Insan Madani.
- Slavin, Robert. 2008. *Cooperative Learning: Teori, Riset, dan Praktik*. Bandung: Penerbit Nusa Media.
- Stronge, HJ (Ed.). 2006. *Evaluating Teaching*. Thousand Oaks, California: Corwin Press Inc.
- Sudijono, A. 1996. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Rajagrafindo Persada.
- Suprijanto. 2007. *Pendidikan Orang Dewasa: dari Teori Hingga Aplikasi*. Jakarta: Bumi Aksara.

## **BAB 2**

### **FUNGSI EVALUASI DALAM PEMBELAJARAN**

#### **PENDAHULUAN**

Penyelenggaraan pembelajaran memerlukan tiga hal yang saling terkait: perumusan tujuan pembelajaran, pelaksanaan pembelajaran, dan tes untuk mengetahui hasil pembelajaran (Gagne dan Briggs, 1974). Dalam tahap perumusan tujuan pembelajaran, diperlukan pemikiran cermat mengenai target yang akan diajarkan, bahan yang akan dipilih, dan tahapan pembelajaran sesuai dengan target yang akan dicapai. Pada tahap pelaksanaan pembelajaran, diperlukan pemilihan metode mengajar yang sesuai dengan tahapan-tahapan bahan yang disusun. Setelah bahan ajar disusun dan proses pembelajaran diterapkan di kelas, guru memerlukan dua informasi penting. Pertama informasi mengenai sejauh mana pembelajar mampu menyerap apa yang sudah diajarkan. Untuk itu, guru memerlukan umpan balik yang di antaranya diperoleh melalui pemberian tes. Selain itu, guru juga memerlukan informasi sejauh mana tujuan pembelajaran, pelaksanaan pembelajaran, dan bahan yang disusun mencapai sasaran sesuai dengan kemampuan pembelajar. Pada tahap ini, guru juga memerlukan informasi yang diperoleh dari hasil tes.

Hakikat tes dalam pembelajaran dengan demikian bukanlah proses tunggal yang terpisah dengan unsur pembelajaran atau berdiri sendiri. Djiwandono (1996) menegaskan, dalam pembelajaran, tes umumnya dikaitkan dengan usaha untuk memperoleh informasi tentang peningkatan kemampuan siswa sebagai hasil pengajaran. Selain itu, tes juga merupakan sumber informasi seperti ketepatan identifikasi dan rumusan tujuan pembelajaran, kesesuaian jenis dan cakupan bahan ajar, kesesuaian metode dan kemampuan mengajar, dan kesesuaian penyediaan waktu. Oleh karena itu, pengajar perlu merancang dan menyediakan tes secara baik dan benar. Pengembangan tes, kesesuaian prosedur, dan penggunaan tes yang sesuai dengan kaidah-kaidah yang berlaku mutlak diperlukan oleh pengajar.

Pengembangan tes sebenarnya juga dipengaruhi oleh metode pembelajaran yang digunakan. Karena itu, pengembangan tes harus juga memperhatikan ciri-ciri metode pembelajaran yang digunakan dan karakteristik bahan ajar yang diberikan. Pemberian tes yang tidak mengacu pada tujuan pembelajaran dan karakteristik bahan ajar, bisa menghasilkan informasi yang menyesatkan. Dalam praktik dewasa ini, banyak pengajar yang menyelenggarakan tes tanpa secara teliti memperhatikan karakteristik bahan ajar dan metode pembelajaran yang digunakan. Tes dalam model ini disusun mengikuti urutan materi yang sudah ada sebelumnya. Misalnya, guru yang mengejar nilai Ujian Nasional umumnya menentukan bahan ajar mengikuti bahan dalam tes. Dengan demikian, proses pembelajaran yang dibuat ditentukan dengan urutan (1) menentukan tes, (2) menyusun bahan ajar, dan (3) memilih metode yang sesuai dengan bentuk tes. Pola ini disebut dengan *backwash*, yang tidak sesuai dengan kaidah pembelajaran. Pembelajaran model *backwash* ini melatih pembelajar menghafalkan konsep dan jawaban soal dan bertentangan dengan tujuan pembelajaran untuk mengembangkan analisis dan berpikir kritis.

Dengan mengacu pada haikat tes dalam pembelajaran yang mencakup tujuan pembelajaran, metode mengajar, dan penyusunan perangkat tes, jelaslah bahwa pengajar dituntut untuk menguasai bukan hanya bagaimana menyusun tes yang sesuai kaidah, tetapi juga harus tahu tentang kurikulum, buku teks, dan metode mengajar. Pemahaman yang kurang proporsional tentang aspek tes dan komponen pendukungnya sebagaimana dikemukakan di atas akan menyebabkan penyelenggaraan tes kurang proporsional dan menyesatkan karena terjadi proses *backwash*.

Dalam pelaksanaan tes secara komprehensif, diperlukan pemahaman mengenai konsep tes secara umum dan seluk-beluk penyusunannya. Selain itu, terkait dengan model pembelajaran di kelas, bentuk tes juga mengalami perkembangan yang nyata sehingga penekanan dan pola pengetesan juga perlu menyesuaikan dengan model tersebut. Makalah ini mencoba mengulas konsep tes

dikaitkan dengan pendekatan pembelajaran, kecenderungan jenis tes, teknik pengembangan tes, dan model tes pada perkembangan terbaru dalam pembelajaran.

## **PERKEMBANGAN ALIRAN DALAM TES**

Secara tradisional tes melibatkan tiga jenis aktivitas: pengukuran, tes, dan evaluasi. Pengukuran mengacu pada identifikasi untuk memperoleh angka-angka sebagai representasi perilaku yang diukur. Dalam tahap ini angka belum memiliki makna. Tes mengacu pada pemberian butir-butir pertanyaan kepada yang dites. Jadi, tes merujuk pada proses pengambilan angka. Selanjutnya, evaluasi artinya proses mempertimbangkan angka-angka hasil pengukuran dibandingkan dengan kriteria tertentu. Dalam evaluasi terjadi pengambilan keputusan sesuai dengan jenis kriterianya, misalnya sangat baik (A), baik (B), cukup (C), kurang (D), dan sangat kurang (E). Dalam praktik, proses memberikan evaluasi itu sering disebut dengan testing walaupun yang dimaksud sebenarnya ialah pengumpulan data berupa angka-angka untuk dibandingkan dengan kriteria sebagai bahan untuk mengambil keputusan. Jadi, yang sebenarnya dilakukan guru untuk mengukur kemampuan pembelajar atau mengetahui kesesuaian proses pembelajaran ialah evaluasi. Proses pengambilan datanya, untuk memudahkan penyebutan, disebut dengan tes atau testing.

Hakikat pengukuran dalam kelas dengan demikian ialah evaluasi. Karena dengan pengukuran saja, pengajar tidak bisa mengambil kesimpulan. Tetapi dengan membandingkan antara hasil pengukuran dengan kriteria, pengajar bisa mengambil keputusan tentang bagaimana keadaan pembelajar sesuai hasil pengukurannya.

Dalam melaksanakan evaluasi sebagaimana dikemukakan di atas, terdapat tiga aliran yang sejauh ini mendominasi sebagaimana dikemukakan oleh Briggs (1994). Ketiga aliran tersebut ialah: (1) *discrepancy model*, (2) *product/outcome oriented*, dan (3) *decision making oriented*. Aliran *discrepancy model* menekankan untuk mengetahui kelemahan dan kekurangan. Pola ini umumnya dilakukan untuk investigasi kasus seperti yang dilakukan oleh polisi atau jaksa. Pada pendekatan

*product oriented*, penekanan evaluasi difokuskan pada hasil akhir. Di sini, proses pembelajaran sebelum pemberian tes tidak penting. Model ini dilaksanakan sekian lama untuk mengetahui hasil belajar siswa dalam praktik pengajaran di Indonesia. Pada pendekatan *decision-making oriented*, hasil evaluasi difungsikan sebagai salah satu bahan pertimbangan untuk membuat keputusan akhir tentang nasib pembelajar. Karena fungsi hasil evaluasi hanya sebagai bagian dalam pengambilan keputusan, berarti diperlukan sumber-sumber lain yang mendukung hasil tes. Dalam pembelajaran, unsur-unsur tersebut merupakan proses dalam pembelajaran di kelas yang terdiri dari kehadiran, tugas, partisipasi, dan pekerjaan lain yang mendukung.

### **PENDEKATAN DALAM TES**

Pelaksanaan tes mengalami perubahan sesuai dengan perkembangan teknologi dan ilmu pengetahuan.

*Pertama*, disebut zaman pra ilmiah. Dalam era ini bentuk tes dan tata cara penilaiannya belum ada standar tertentu. Begitu juga syarat-syarat pengembangan tes yang baik, belum ditetapkan dan belum disepakati. Bahan yang diteskan juga belum standar. Umumnya bahan tes mengikuti kemauan apa yang dibuat guru. Materi yang dibuat mengacu pada penguasaan konsep bidang ilmu tertentu sampai sangat mendalam. Ketentuan lulus tidaknya pembelajar dalam menjawab tes sangat tergantung pada *mood* si pengajar. Karena tidak ada kriteria yang digunakan, guru menjadi sangat *powerful*. Apa yang akan diteskan tidak bisa dilacak dari bahan ajar atau metode mengajar yang pernah disampaikan oleh guru di kelas.

*Kedua*, era psikometrik atau era ilmiah. Era ini disebut ilmiah karena pemberian tes sudah didasarkan pada kriteria dan kaidah tertentu. Kriteria yang digunakan merujuk pada penggunaan angka-angka untuk merepresentasikan perilaku belajar. Jadi, proses tes diberikan dengan mengidentifikasi perilaku belajar ke dalam angka-angka. Proses penilaian ini sejalan dengan perkembangan psikometrik. Aliran ini memandang bahwa segala sesuatu yang abstrak harus bisa dikuantifikasi. Data

yang bukan berupa angka dianggap lemah, sehingga data-data tersebut harus diubah dulu ke dalam angka-angka.

Kehadiran aliran psikometrik ini memang menandai revolusi ilmu pengetahuan. Patokan penilaian mulai bisa digunakan secara baku dan seragam. Salah satu contoh penggunaan kriteria ialah penggunaan kategori sangat baik (80-100), baik (70-79), sedang (60-69), kurang (50-59), dan sangat kurang (10-49). Yang menonjol dalam pendekatan ini ialah kategori dalam "slogan normal". Menurut aliran ini, segala sesuatu di dunia terjadi secara normal. Karena itu, hasil tes juga dikelompokkan ke dalam jenis normal. Butir-butir pertanyaan yang disusun juga harus mengacu pada "hukum serba normal ini".

Implikasi dari hukum serba normal ini ialah: peserta tes rata-rata "hanya boleh" mendapat nilai normal (sedang atau dalam rentangan 60). Siswa yang sangat pintar, pintar, atau sangat kurang hanya berjumlah sedikit saja. Inilah sebabnya, pada era ini misalnya dari 100 mahasiswa yang mengikuti ujian, antara 70% memperoleh nilai C. Kalau ada yang mendapat A, B, D, atau E hanya antara 10% saja. Dalam beberapa kasus malahan rentangan nilai mahasiswa hanya C, D dan E.

Implikasi lain di bidang pengembangan tes juga substansial. Dalam era ini sudah disepakati aturan mengenai kualitas butir soal, validitas, dan reliabilitas. Kualitas butir soal mengacu pada pembobotan tingkat kesulitan, daya pembeda, dan penggunaan distraktor yang efektif. Penggunaan kriteria validitas dan reliabilitas mengacu pada pembuatan kisi-kisi tes dan perbandingan hasil tes yang sedang disusun dengan hasil tes lain yang sudah dianggap memiliki kualitas yang baik.

Dari segi bahan ajar dan metode mengajar, era ilmiah ini menekankan pada pembelajaran berpusat pada siswa. Karena itu, tujuan pembelajaran dan tujuan tes ditekankan pada penguasaan konsep tentang materi yang diajarkan sebanyak-banyaknya. Analisis tingkat kesulitan butir soal mengacu pada taksonomi pembelajaran yang dikemukakan oleh Bloom (1959). Tujuan pembelajaran dikelompokkan ke dalam tiga jenis: kognitif, afektif, dan psikomotorik, masing-masing dielaborasi menggunakan kata-kata kerja operasional yang sudah baku.

Dalam tahap awal, taksonomi yang banyak diikuti ialah aspek kognitif. Untuk itu, ukuran tingkat kesulitan butir soal menggunakan ranah: *recall, comprehension, application, analysis, synthesis, dan evaluation*. Domain afektif dan psikomotorik kurang digunakan.

**Ketiga**, pendekatan proses. Pendekatan proses dalam tes mengacu pada model evaluasi yang menekankan pada penggunaan bahan ajar dalam kehidupan senyatanya di lapangan. Pendekatan ini dalam pembelajaran sains merujuk pada belajar inquiry atau pendekatan proses. Dalam pembelajaran bahasa, pendekatan ini merujuk pada *communicative approach*. Hakikat evaluasi ialah untuk mengetahui kemampuan siswa dalam menerapkan konsep yang dipelajari secara nyata di lapangan.

Pada era ini, penggunaan jenis tes objektif dan esai sangat populer. Tes objektif digunakan untuk menguji kompetensi dan tes esai digunakan untuk menguji penalaran. Di luar negeri, ranah yang dinilai sudah berkembang tidak hanya pada aspek kognitif tetapi juga afektif dan psikomotorik. Di Indonesia terjadi kecenderungan menggunakan aliran tes *product oriented*. Untuk itu, model tes yang lebih ditekankan ialah tes objektif karena dianggap mudah dikoreksi. Selain itu, untuk tujuan penggunaan tes massal seperti EBTANAS atau UMPTN, tes objektif mudah dikoreksi dan dengan cepat memberikan informasi.

Kelemahan era ini secara nyata melahirkan lulusan yang kurang daya nalar dan analisisnya. Penguasaan logika kurang tajam dan argumentasi dalam menyampaikan pandangan tertentu tidak kritis. Hal ini juga secara langsung berdampak pada proses pembelajaran di kelas. Dalam pembelajaran di kelas, tujuan utamanya ialah menuntaskan materi. Karena materi akan diukur menggunakan tes nasional, apabila bahan tidak habis akan berdampak pada nasib si guru.

**Keempat**, pendekatan berbasis konteks. Dalam pendekatan ini, yang diutamakan dalam evaluasi ialah proses bagaimana pembelajar mencapai kompetensi tertentu. Untuk itu, penilaian diarahkan pada portofolio pembelajar. Portofolio ini menuntut adanya rekam jejak hasil belajar, seperti PR, partisipasi, kehadiran, tugas,

dan juga hasil tes harian. Dalam pendekatan berbasis konteks, tujuan tes tidak semata-mata didasarkan pada apa yang diajarkan saja, tetapi meliputi: standar kompetensi, kompetensi dasar, dan kompetensi minimal.

Dalam kompetensi dasar materi ajar sudah ditetapkan dalam kurikulum dan kemampuan dasarnya dibandingkan dengan kriteria nasional. Misalnya, dalam pembelajaran speaking, kompetensi dasarnya ialah bisa berkomunikasi dengan *native speaker*. Jadi, setelah speaking diajarkan, mahasiswa harus mampu berkomunikasi dengan *native speaker*.

Tapi tidak semua mahasiswa akan bisa mencapai tingkatan itu. Karena itu, dosen harus membuat tahapan materi dan proses pembelajaran secara cermat. Yang pertama-tama dilakukan ialah menetapkan kompetensi dasar mahasiswa. Kompetensi dasar ini sesuai dengan bahan yang akan diajarkan. Misalnya, agar mampu berkomunikasi secara lancar, mahasiswa harus menguasai topik secara instan, tidak boleh mempersiapkan diri tertulis atau diskusi, dan kalimat-kalimat serta ucapan harus disampaikan secara jelas. Inilah kompetensi dasar yang menjadi target pembelajaran dosen. Setelah beberapa kali dilatih dengan target kompetensi dasar yang demikian tadi, mahasiswa akan mampu mencapai standar kompetensi.

Karena kemampuan mahasiswa di kelas tidak merata, tentu ada yang mencapai maksimum dan ada yang kurang. Untuk itu, dosen memerlukan kompetensi minimal. Ini disebut dengan standar kompetensi minimal. Dengan demikian, mahasiswa yang sudah menempuh speaking dari dosen A, sekurang-kurangnya bisa berkomunikasi menggunakan kalimat yang benar dan ucapan yang jelas.

Implikasi dari pendekatan berbasis konteks ini cukup kompleks. Pertama, ranah yang diukur tidak lagi terbatas pada aspek kognitif, tetapi juga afektif dan psikomotorik. Dengan demikian, tataran berpikir dalam menjawab pertanyaan bukan lagi ditekankan pada kemampuan mengingat atau menghafal, tetapi sekurang-kurangnya menganalisis dan mengevaluasi. Kompetensi dalam konteks ini mengacu pada penguasaan skill sekaligus: mengetahui konsep, bisa menjelaskan, dan bisa

mempraktikkan sewaktu-waktu diminta. Kedua, tujuan pembelajaran, model pembelajaran, dan pola evaluasi yang digunakan tidak cukup hanya tes objektif, tetapi juga harus esai dan praktik. Dalam hal ini, secara umum digunakan portfolio, proyek, performance, dan menulis paper. Implikasinya, selain pengajar harus menguasai model tes yang sudah lazim, mereka harus dilatih dan dibiasakan menggunakan tes berbasis kelas dalam model pembelajaran berbasis konteks.

### **PERMASALAHAN DALAM EVALUASI**

Dengan merunut perkembangan tes dan karakteristiknya yang demikian ini, permasalahan yang muncul ialah: bagaimana menguasai jenis tes dan mengembangkan jenis tes sesuai dengan tujuan pembelajaran dan perkembangan metode mengajar? Uraian ini mencoba meringkas lima hal dalam upaya menjawab pertanyaan tersebut. Kelima hal itu ialah: jenis tes, format tes, syarat tes yang baik, interpretasi hasil tes, pengembangan butir tes, dan evaluasi berbasis kelas.

#### **Jenis Tes**

Jenis tes di sini dibatasi hanya pada jenis tes berdasarkan tujuan penyelenggaraannya dan jenis tes berdasarkan cara membuatnya. Berdasarkan tujuan penyelenggaraannya, tes dibagi ke dalam tiga jenis: tes formatif, tes sumatif, dan tes profisiensi. Tes formatif atau tes harian diselenggarakan setelah satu atau dua unit selesai diajarkan. Penekanan pada tes ini ialah ketuntasan bahan ajar dalam waktu pendek. Di perguruan tinggi, tes formatif biasanya diberikan untuk memberi penguatan penguasaan bahan ajar dan variasi tes diberikan dalam bentuk kuiz, tugas, atau UTS (ujian tengah semester). Tes sumatif diselenggarakan dalam rentang waktu setelah satu periode belajar diselesaikan, misalnya satu semester. Karena akhir semester bisa juga terjadi bersamaan dengan kenaikan kelas atau lulusan, tes sumatif difungsikan juga untuk tes kenaikan kelas atau evaluasi akhir. Tes formatif dan sumatif mengacu pada ketuntasan proses belajar dalam periode tertentu. Tes lain yang tidak terikat oleh waktu pelaksanaan pembelajaran, tetapi mengacu pada hasil

yang bisa ditunjukkan pembelajar disebut dengan *achievement test*. Tes jenis ini disebut juga tes profisiensi atau *mastery test*. Jika tes formatif dan sumatif menekankan pada penguasaan bahan yang sudah diajarkan pada kurun waktu tertentu, tes profisiensi mengacu pada sejauh mana lulusan tertentu menguasai pengetahuan tertentu setelah mereka lulus dari tingkat tertentu.

Dilihat dari cara membuatnya, tes dibedakan ke dalam tes buatan sendiri dan tes terstandar. Tes buatan sendiri (sering disebut *teacher-made test*) ialah tes yang dibuat sendiri oleh pengajar untuk mengetahui keberhasilan pembelajaran dan kemampuan siswanya. Disebut tes buatan sendiri karena tes tersebut akan digunakan sendiri di kelas. Karena itu, tes buatan sendiri disebut juga tes untuk evaluasi di kelas (*classroom test*) dan digunakan untuk tes formatif.

Implikasi dari tes ini ialah: guru tidak perlu menyusun kisi-kisi secara rinci sebelum menyusun butir soal. Selain itu, tes tidak perlu diuji-coba untuk mengetahui kualitas butirnya sebelum digunakan untuk mengetes siswa. Dengan kriteria demikian, tes buatan guru biasanya kualitasnya rendah karena tidak melalui seleksi butir soal yang teliti. Selain itu, bentuk soal yang dibuat juga cenderung yang memudahkan pembuat, misalnya berbentuk esai dan disusun untuk pokok bahasan terbatas.

Tes terstandar ialah tes yang kualitasnya memenuhi kriteria tertentu. Tes ini disusun secara cermat melalui tahapan yang jelas dan proses seleksi butir soalnya dilakukan melalui uji-coba (*try-out*). Yang dimaksud disusun secara cermat mencakup kegiatan menyiapkan kisi-kisi, proporsi materi diteliti secara mendalam, perumusan butir pertanyaan dibuat secara rinci, dan setelah butir-butir disusun, dilakukan uji-coba untuk menyeleksi butir-butir soal yang masuk kriteria. Karena prosesnya yang demikian, tes terstandar disusun dalam waktu yang tidak pendek.

Jika tes buatan sendiri digunakan untuk tes kelas saja, tes terstandar digunakan untuk tujuan massal, misalnya UNAS atau UMPTN. Kekuatan tes terstandar ialah adanya kalibrasi butir soal. Kalibrasi ini bertujuan untuk mengetahui tingkat kesulitan butir, daya pembeda, efektivitas distraktor, validitas, dan reliabilitas.

### Format Tes

Format tes ialah bentuk tes dilihat berdasarkan cara memberikan skor. Format tes ada dua: objektif dan non-objektif. Tes objektif ialah tes yang caranya menskor diberikan secara diskrit, jawaban benar mendapat poin dan jawaban salah mendapat poin nol. Karena sifat pemberian skor didasarkan pada fakta benar salah, tes ini disebut dengan tes objektif. Jenis tes objektif ini antara lain: menjodohkan, benar salah, pilihan ganda, isian pendek dengan satu jawaban.

Format tes yang kedua disebut dengan tes non-objektif. Skoring untuk tes jenis ini diberikan melalui skala (scaling points) dari rentangan minimum ke maksimum. Jawaban yang memenuhi kriteria penuh mendapat skor semakin tinggi dan yang kurang penuh mendapat skor kurang. Penskoran untuk tes jenis ini, melibatkan kesan dan perasaan penilai. Selain itu, jawaban tes berupa uraian yang menunjukkan penalaran si penjawab. Karena cirinya yang demikian, tes ini disebut juga tes subjektif atau tes esai. Kekuatan tes esai ialah tes ini menguji penalaran, analisis, sintesis dan evaluasi. Semakin mendalam uraian dan didukung dengan data-data, semakin sempurna jawaban tes. Dalam bentuk lebih luas, tes jenis ini bisa berupa paper, laporan hasil penelitian, atau kajian terhadap topik tertentu. Berikut ialah contoh tes esai.

- (1) Sebutkan tiga alasan mengapa bangsa Indonesia memilih negara demokrasi (skor 15).

Butir pertanyaan dalam soal dalam bentuk kalimat disebut dengan stem, di dalam stem terdapat task (beban pekerjaan) yang harus dijawab oleh peserta tes. Pada contoh (1) di atas, pertanyaan yang harus dijawab ialah **negara demokrasi**, sedangkan task yang harus dikemukakan penjawab ialah **tiga alasan**.

Penilaian yang diberikan untuk jawaban soal di atas, harus didasarkan pada kriteria dengan rentangan skor tertentu. Ilustrasi kriteria dan rentangan skornya bisa dikemukakan berikut ini:

Tabel 1. Kunci Soal Esai dan Rentangan Skor

No	Indikator Jawaban	Baik	Sedang	Kurang
1	Indonesia terkungkung orde baru	5	3	1
2	Dasar negara Indonesia ialah demokrasi	5	3	1
3	Indonesia ingin lepas dari kekangan militer	5	3	1

Tabel di atas menjelaskan bahwa kunci jawaban soal terdiri dari tiga pokok pikiran. Setiap pokok pikiran diberi poin 5, sehingga apabila satu jawaban saja yang benar, skor yang diberikan 5, 10 untuk 2 jawaban benar, dan 15 untuk tiga jawaban benar. Masalahnya ialah, dalam menyajikan uraian tidak semua penjawab menggunakan kalimat yang sama. Dalam hal ini, guru bisa menempuh dua cara: jawaban dianggap benar apabila poin jawaban sama dengan kunci atau uraian dan poin jawaban harus mendetil. Angka 15 di belakang pertanyaan menunjukkan berapa skor maksimum jika jawaban benar. Angka ini memudahkan pengoreksi untuk menentukan skala penilaian dan memberi petunjuk pada penjawab untuk membuat seberapa mendalam uraian yang diperlukan.

### Syarat Tes Yang Baik

Berbagai kepustakaan menyebutkan ada tiga syarat tes yang baik: kepraktisan, validitas, dan reliabilitas. Karena tes yang baik juga melibatkan pemilihan tingkat kesulitan butir yang baik, penulis berpendapat, kualitas butir soal juga bagian dari syarat tes yang baik. Jadi secara lengkap, syarat tes yang baik meliputi: tingkat kesulitan butir soal, validitas, dan reliabilitas. Uraian mengenai tingkat butir soal sudah dibahas di muka dengan mengacu pada uji-coba butir soal.

Kepraktisan ialah kemudahan dalam menggunakan butir tes. Kemudahan ini meliputi kemudahan dalam skoring dan penggunaan. Tes yang didesign untuk dikoreksi menggunakan komputer, mungkin praktis dan mudah apabila digunakan di dalam kota yang ada listrik. Tetapi tes tersebut akan menyulitkan apabila digunakan

di perbatasan Indonesia-Malaysia yang belum ada listrik. Karena kepraktisan tidak terkait dengan substansi tes, beberapa ahli menganggap kepraktisan bukan syarat tes yang baik.

Syarat kedua ialah validitas. Validitas artinya kecocokan antara hasil tes dengan kemampuan yang diuji. Misalnya, siswa yang mendapat nilai membaca Al Quran 9, harus menunjukkan kemampuan melafalkan ayat-ayat dengan sempurna. Apabila misalnya, siswa tadi disuruh membaca Alfatikah saja tidak lancar, berarti skor 9 tadi tidak cocok. Berarti hasil tes tadi tidak valid. Jadi yang dimaksud valid ialah skornya, bukan tesnya.

Pendekatan validitas digunakan untuk menunjukkan bahwa hasil tes memiliki kecocokan dengan penampilan yang diuji. Validitas terdiri dari dua jenis: validitas logis dan validitas empiris. Validitas logis ialah cara menunjukkan hasil tes cocok berdasarkan logika atau alasan. Validitas logis terdiri dari: *face validity*, *content validity*, dan *construct validity*. *Face validity* ialah kecocokan dilihat dari penampilan tes. Misalnya, guru yang menguji kemampuan menulis, di dalam kelas ternyata siswa disuruh berbicara. Dalam melihat aktivitas ini, kita bisa melihat penampakkannya dengan berkomentar: “Tes menulis kok bicara ya?”.

Validitas isi (*content validity*) ialah pembuktian melalui kecocokan antara butir tes dengan tujuan tes dan sebaran bahan tes. Validitas ini bisa dilihat dari kisi-kisi soal. Validitas konstruk ialah pencocokan butir tes dengan teori belajar yang mendasari. Contoh validitas konstruk ialah sebagai berikut. Kelas A diajar oleh orang Jepang untuk membuat radio. Kelas B diajar insinyur ITB dalam membuat radio. Baik orang Jepang maupun insinyur ITB telah menampilkan perilaku belajar, metode, dan langkah-langkah pembelajaran sesuai dengan konsep dan teori yang dianutnya. Masing-masing menerapkan konstruk yang berbeda. Apabila kelas A dan kelas B di akhir pembelajaran sama-sama mampu membuat radio, berarti konstruk yang digunakan oleh orang Jepang dan Insinyur ITB tadi memiliki kesamaan. Valid.

Validitas jenis yang kedua ialah validitas empiris. Validitas ini terdiri dari validitas prediktif dan validitas kriteria. Kedua jenis validitas ini memerlukan nilai

pembandingan yang diperoleh dari tes lain yang sejenis. Proses perbandingan dilakukan dengan cara uji statistik. Jika nilai uji statistik menunjukkan korelasi yang tinggi, berarti validitas skor tes yang dicari memiliki kesamaan dengan validitas skor tes pembandingan.

Syarat ketiga ialah reliabilitas. Reliabilitas artinya keajegan atau konsistensi. Yang dimaksud konsisten ialah skor yang dihasilkan oleh tes. Andaikan seperangkat tes dikoreksi oleh penilai A dan penilai B, skor yang diberikan oleh kedua penilai tersebut harus kurang lebih sama, konsisten. Jika terjadi perbedaan skor yang mencolok, berarti skor tersebut tidak konsisten atau tidak reliabel.

Kalau validitas bisa dibuktikan melalui logika dan uji statistik, reliabilitas sepenuhnya harus menggunakan uji statistik. Reliabilitas bisa diperoleh melalui empat cara: (1) *test-retest*, (2) *split-halves method*, (3) *equivalent forms*, dan (4) metode Kuder-Richardson. Baik validitas dan reliabilitas memerlukan hitungan statistik sederhana utamanya untuk melihat korelasi.

### **Interpretasi Hasil Tes**

Hasil tes berupa skor selanjutnya harus diinterpretasi dengan cara dibandingkan dengan kriteria tertentu. Ada dua cara dalam menginterpretasi hasil tes: menggunakan Patokan Acuan Norma (PAN) dan Patokan Acuan Patokan (PAP). Dalam PAN, nilai batas lulus (*passing grade*) tidak ditentukan lebih dulu. Patokan yang digunakan ialah rata-rata kelas. Jadi, nilai rata-rata dikategorikan sebagai nilai sedang atau C. Selanjutnya, sebaran nilai ke dalam kategori A, B, D dan E ditetapkan berdasarkan besaran standar deviasai ditambah rata-rata. Dalam penilaian PAP, kriteria batas lulus sudah ditetapkan secara ketat, misalnya 70 atau 80. Dalam praktik, penggunaan PAN dan PAP sering dikombinasikan sehingga melahirkan kriteria tertentu. Beberapa kriteria yang digunakan di perguruan tinggi di Indonesia dan sekolah internasional disajikan dalam tabel berikut.

Tabel 2. Contoh Kriteria PAP

No	Skala 100	Skala 10	Skala 4	Nilai	Kualifikasi
1	85-100	8.5-10	4	A	Sangat baik
2	70-84	7.0-8.4	3	B	Baik
3	55-69	5.5-6.9	2	C	Sedang
4	50-54	5.0-5.4	1	D	Kurang
5	10-49	1.0-4.9	0	E	Sangat Kurang

Tabel 3. Kombinasi PAN dan PAP

No	Rentangan	Skor	Nilai	Kualifikasi
1	80-100	4	A	Sangat baik
2	70-79	3	B	Baik
3	60-69	2	C	Sedang
4	50-59	1	D	Kurang
5	10-49	0	E	Sangat Kurang

Tabel 4. Kriteria Penilaian di Perguruan Tinggi di Indonesia

No	Skala 100	Skala 10	Skala 4	Nilai	Kualifikasi
1	91-100	3.8-4.0	4,0	A	Sangat baik
2	86-90	3.6-3.79	3,7	A-	
3	81-85	3.2-3.59	3.3	B+	Baik
4	76-80	2.9-3.19	3.0	B	
5	70-75	2.5-2.89	2.7	B-	Sedang
6	60-69	1.5-2.49	2.0	C	Kurang
7	50-59	0.5-1.49	1.0	D	Sangat Kurang
8	10-49	0.0-0.49	0	E	

Tabel 5. Kriteria Penilaian di Sekolah Internasional

No	Skala 100	Skala 4	Nilai	Kualifikasi
1	97-100	4,3	A+	Outstanding
2	93-96	4.0	A-	
3	90-92	3.7	A	Sangat baik
4	87-89	3.3	B+	
5	83-86	3.0	B	
6	80-82	2.7	B-	Memuaskan
7	77-79	2.3	C+	
8	73-76	2.0	C	
9	70-72	1.7	C-	Di bawah rata-rata
10	67-69	1.3	D+	
11	63-66	1.0	D	
12	60-62	0.7	D-	Gagal
13	59/below	0	F	

Tabel 6. Kriteria Penilaian TOEFL

No	Pencil Based	Computer Based	Kualifikasi
1	650 up	280 up	Dijamin diterima di semua program
2	600-649	250-279	Dijamin diterima untuk program S1
3	550-599	213-249	Bisa diterima di program S2
4	500-549	173-212	Bisa diterima di program S1
5	450-499	133-172	Perlu review pribadi
6	449-less	132-less	Harus ikut pelatihan khusus

### Authentic Assessment

Perkembangan model pembelajaran berbasis konteks menuntut model penilaian yang menekankan proses. Evaluasi didasarkan pada proses pembelajaran dalam kelas yang terjadi berkesinambungan dan saling melengkapi. Evaluasi demikian ini disebut dengan *authentic assesment* atau penilaian berbasis kelas (*classroom based evaluation*). *Authentic assesment* terdiri dari empat jenis: *portfolio*, *project*, *performance*, dan *extended writing*.

Penilaian portfolio merupakan rekam jejak kemampuan akademis selama mengikuti pembelajaran. Portfolio berisi artifak atau bukti-bukti pekerjaan yang dilakukan siswa selama periode tertentu, misalnya catatan, tugas, pekerjaan rumah. Portfolio bisa berupa kliping yang berisi kumpulan dokumen pekerjaan siswa dalam mengikuti kegiatan pembelajaran.

*Project* ialah pekerjaan mandiri yang harus dikerjakan siswa baik secara individu maupun kelompok. Proyek bertujuan untuk memberikan beberapa tujuan pembelajaran sekaligus dan harus dikuasai oleh siswa melalui proses: persiapan, perencanaan, mengerjakan, memecahkan masalah, menyusun laporan, dan menyampaikan hasil.

*Performance* ialah penampilan siswa terhadap suatu kinerja. Penampilan ini bisa berupa menyajikan ide dalam seminar, bermain drama, menyajikan karya ilmiah, menyanyi, atau melakukan tugas tertentu. Dalam menilai performance, siswa lain bisa diminta sebagai penilai.

*Extended writing* ialah menulis dalam tingkat lanjut. Menulis jenis ini bisa berupa paper, mini riset, hasil observasi, kajian buku, menyusun karya tulis tertentu, atau menyusun laporan penelitian.

Dalam penilaian autentik, yang penting diperhatikan ialah model penilaian dan penguasaan kriteria penilaian. Penilaian autentik ini melibatkan pola berpikir tahap tinggi dan berbagai tujuan belajar yang kompleks menjadi satu. Untuk itu, guru perlu menguasai dengan baik kaidah penilaian esai, karya tulis, kinerja dan kegiatan langsung yang menampilkan kemampuan belajar sekaligus.

## **PENUTUP**

Evaluasi merupakan syarat mutlak dalam mengetahui umpan balik hasil pembelajaran dan daya serap pembelajar terhadap materi yang diajarkan. Model dan teknik evaluasi berkembang sesuai dengan perubahan model pembelajaran di kelas. Untuk menguasai evaluasi dengan baik, pengajar perlu menguasai juga kaidah, norma, prosedur, dan aturan dalam pembuatan tes, menyeleksi butir tes yang baik,

dan menginterpretasi hasil tes. Syarat tes yang baik yang meliputi: seleksi butir soal, validitas, dan reliabilitas merupakan hal yang tidak bisa dihindari. Dalam menginterpretasi hasil tes, digunakan PAN dan PAP. Masalahnya, standar lulus yang digunakan di Indonesia selalu lebih rendah. Dalam mengkaji tes untuk pembelajaran, aliran yang sekarang cocok dengan pendekatan berbasis konteks ialah authentic assessment. Selain itu penggunaan tes jenis esai juga terbukti mampu meningkatkan penalaran siswa secara lebih baik.

#### **DAFTAR RUJUKAN**

- Anastasi, A. 1982. *Psychological Testing*. New York: Macmillan Publishing Co.
- Budiharso, T. 2005. *Evaluasi Berbasis Kelas dalam Pembelajaran Bahasa Kedua*. Jakarta: Lekdis.
- Djiwandono, M.S. 1996. *Tes Bahasa dalam Pengajaran*. Bandung: Penerbit ITB Bandung.
- Ebel, R. 1979. *Essentials of Educational Measurement*. New Jersey: Prentice Hall Inc.
- Grondlund, N. 1985. *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Co.
- Johnson, E.B. 2002. *Contextual Teaching and Learning: What It is and Why It is to Stay*. Thousand Oaks, California: Corwin Press Inc.
- Lion, A. 1980. *Measurement and Evaluation of Learning*. Iowa: Brown Company Publisher.

## **BAB 3**

### **TES BUATAN GURU DAN TES TERSTANDAR**

#### **DEFINISI**

Tes buatan guru menurut Harris (1969:1) ialah tes yang disiapkan, dilaksanakan, dan diskor oleh seorang guru untuk mengukur pencapaian siswanya sendiri. Tujuan pengetesan ialah tujuan pengajaran, sedangkan materi tes diambil dari bahan pengajaran juga. Tes buatan guru memiliki ketepatan materi tes dan tujuan apa yang sudah diberikan di kelas. Tes buatan guru digunakan untuk mengukur sejauh mana kemajuan siswa dan mendiagnosa kekuatan dan kelemahan siswa dalam penguasaan materi tertentu.

Pada umumnya, pengertian tes buatan guru selalu dihubungkan dengan tes terstandar. Tes terstandar menurut Ebel (1986:267) ialah tes yang (1) tes yang disusun oleh para ahli penyusun tes dan telah diujicobakan, dianalisis, dan diperbaiki, (2) memiliki petunjuk pelaksanaan dan pemberian skor yang jelas, dan (3) memiliki acuan norma untuk penginterpretasian skor. Tes terstandar dapat dipakai berulang-ulang di berbagai sekolah bahkan di berbagai negara.

Dengan membandingkan dua definisi di atas, bisa disimpulkan bahwa tes yang membedakan secara tegas antara tes buatan guru dengan tes terstandar terutama pada proses ujicoba dan analisis butir soalnya. Tes buatan guru karena digunakan untuk kalangan sendiri, sering tidak memerlukan ujicoba. Sebaliknya, tes terstandar memerlukan ujicoba sebagai kriteria utama untuk menyeleksi butir soal yang baik.

Karena tidak diujicobakan, tes buatan guru tidak dilengkapi informasi mengenai kualitas butir, validitas dan reliabilitas soal. Begitu juga sistem skoringnya, sering tergantung pada pola yang diterapkan guru sendiri. Dengan demikian, walaupun para guru bergabung untuk menyusun soal, tetapi apabila proses ujicoba, revisi soal, dan analisis butir belum dilaksanakan, tes tersebut masih terkatagori pada tes buatan guru.

## **CIRI-CIRI TES YANG BAIK**

Menurut Harris (1969:13) dan Djiwandono (1996:90) sebuah tes dikatakan baik apabila tes tersebut memenuhi persyaratan utama yaitu: valid, reliabel dan praktis. Tes yang baik merupakan tuntutan pokok dalam mengetahui hasil belajar bahasa. Menurut Djiwandono (1996:90) melalui penyelenggaraan tes, diharapkan dapat diperoleh informasi mengenai berbagai segi penyelenggaraan pengajaran bahasa. Di samping hasil informasi mengenai hasil belajar siswa sebagai salah satu segi yang terpenting, penyelenggaraan tes dalam pengajaran bahasa memberikan pula informasi tentang ketepatan identifikasi tujuan, kesesuaian bahan pengajaran, kesesuaian dan keefektifan metode pengajaran, cukup tidaknya latihan yang diberikan, kesulitan belajar siswa. Semua informasi itu dapat diperoleh atas dasar pengamatan terhadap tingkat keberhasilan siswa seperti tercermin pada nilai-nilai yang dicapai, dan kajian terhadap tingkat dan jenis kesalahan yang dibuat oleh peserta tes. Mengingat begitu pentingnya fungsi tes, maka betapa perlunya mengusahakan agar tes yang digunakan dalam pengajaran benar-benar merupakan alat yang baik dan dapat diandalkan.

### **1. Validitas**

Validitas suatu tes diartikan sebagai kesesuaian antara skor yang diperoleh siswa dengan tujuan yang hendak diuji (Gronlund, 1979; Ebel, 1979; Djiwandono, 1996). Dengan demikian, yang dimaksud dengan valid bukanlah tesnya, tetapi skor yang diperoleh siswa sudah sesuai dengan tujuan tes atau belum. Andaikan soal yang ditulis dimaksudkan untuk menguji kemampuan membaca, maka butir-butir yang ditulis untuk soal harus menanyakan pemahaman isi wacana. Jika ada pertanyaan yang mengharuskan penjawab soal mengenai menulis, maka butir soal tersebut tidak akan menunjukkan hasil pemahaman. Oleh sebab itu, skor yang diperoleh siswa tidak menunjukkan kemampuan yang sebenarnya mengenai membaca.

Secara umum, validitas dikategorikan ke dalam dua jenis: validitas logis dan validitas empiris (Gronlund, 1979). Validitas logis, menurut Djiwandono (1996) diperoleh apabila unsur-unsur pembuktiannya memerlukan penalaran. Sebaliknya, validitas empiris memerlukan pembuktian melalui tes statistik dan memerlukan adanya kriteria sebagai pembanding.

Jenis-jenis validitas, karena itu bisa dibedakan menjadi empat jenis: validitas tampak luar, validitas isi, validitas kriteria, validitas konstruk dan validitas (Djiwandono, 1996; Gronlund, 1979; Oller, 1979). Se jauh mana sebuah tes memiliki tingkat validitas yang tinggi, diukur dengan seberapa banyak evidence yang bisa digunakan untuk mendukung kualitas tes. Misalnya, sebuah tes yang memiliki evidence isi, kriteria dan konstruk, akan lebih meyakinkan dibanding tes yang hanya memiliki evidence dalam hal validitas kriteria saja.

Proses kemandapan evidence untuk mendukung validitas dimulai ketika butir-butir soal dirancang dalam bentuk kisi-kisi (*Table of Specification*) yang berisi proporsi tingkat kesulitan dan penyebaran materi yang akan diuji, penulisan butir soal dan hasil try-out untuk mengetahui tingkat kesulitan dan daya beda setiap butir soal. Apabila soal tersebut berbentuk pilihan ganda, dua hal tersebut masih harus ditambah dengan efektivitas option (*distractor*).

Apabila validasi suatu tes dilakukan menggunakan uji statistik, tingkat koefisien validitas berkisar antara 0-1. Semakin mendekati angka 1, hasil tes semakin valid. Ebel (1979) memberikan angka .50 untuk tes yang digunakan di kelas dan .90 ke atas untuk tes terstandar.

Harris (1969:19) menjelaskan bahwa tes dikatakan valid apabila tes tersebut melakukan dua hal secara baik: (1) mengukur secara tepat apa yang hendak diukur dan tidak mengukur hal lain yang bukan menjadi tujuannya, dan (2) seberapa baik tes tersebut mengukur tujuan tersebut. Menurut Djiwandono (1996:91) validitas merupakan ciri tes yang amat penting, yang seharusnya dimiliki oleh setiap tes yang digunakan

untuk berbagai tujuan, termasuk tujuan pengajaran bahasa. Secara konvensional, validitas diartikan sebagai ciri yang menunjukkan adanya kesesuaian antara tes dengan apa yang ingin diukur dengan menggunakan tes itu. Secara lebih tepat, ditegaskan oleh Djiwandono, bahwa sebenarnya validitas bukan merupakan ciri yang terkait pada tesnya sebagai alat, tetapi apakah nilai yang diperoleh peserta tes bisa menunjukkan kemampuan mengenai bahan yang diujikan.

Dengan demikian, validitas tes tergantung pada beberapa hal. *Pertama*, validitas tergantung pada tujuan pengadaan tes. Untuk apa tes itu dilakukan; untuk menempatkan siswa, menjaring siswa, melihat kemampuan siswa. *Kedua*, validitas tes tergantung pada kelompok yang dites. Suatu tes yang valid untuk anak-anak, akan menjadi tidak valid untuk orang dewasa. *Ketiga*, validitas tes tergantung juga pada situasi dan kondisi saat tes itu digunakan.

Dewi (1991:26-30) membedakan validitas ke dalam dua aspek: aspek kuantitatif dan aspek kualitatif. Secara garis besar, aspek kuantitatif digolongkan ke dalam validitas empiris (*empirical validity*), sedangkan aspek kualitatif dikelompokkan ke dalam validitas logis (*logical validity*).

### **Validitas Empiris**

Validitas empiris ialah validitas suatu tes yang dicapai dengan cara membandingkan atau mengkorelasikan skor tes dengan skor tes lain yang dipakai sebagai kriteria. Yang termasuk dalam validitas empiris ini ialah (1) validitas peramalan (*predictive validity*), dan (2) validitas bandingan (*concurrent validity*). Validitas peramalan menunjukkan apakah tes yang dikerjakan sekarang dapat meramalkan performansi seseorang di masa yang akan datang (Tuckman, 1975:238). Misalnya, suatu sekolah mengadakan dua jenis tes, yaitu tes bakat membaca yang diteskan pada semester 2, dan tes membaca pemahaman yang diteskan pada semester ke 4. Setelah skor kedua tes dikorelasikan, diketahui bahwa hasilnya 0.58. Angka ini menunjukkan bahwa tes bakat mempunyai validitas

ramalan untuk membaca pemahaman karena tes tersebut bisa meramalkan kemampuan siswa di bidang membaca.

Satu lagi validitas empiris ialah validitas bandingan. Validitas bandingan ialah validitas tes ditinjau dari korelasinya dengan dengan kemampuan yang dimiliki saat ini. Kemampuan (skor kemampuan) yang ada pada saat tes dilaksanakan dapat diperoleh dari nilai dalam rapor semester saat itu, dari hasil tes lain yang sejenis yang memiliki validitas, atau dari guru yang membina siswa tersebut.

Dengan demikian, validitas logis pembuktiannya memerlukan alat berupa tes. Dua metode yang sering digunakan untuk menguji validitas empiris ini di antaranya: metode *rank-difference* untuk sejumlah skor kecil; dan metode product moment untuk sejumlah skor besar. Nilai validitas ini dinyatakan dengan koefisien dan besarnya berkisar antara 0 hingga 1 (Djiwandono, 1996:95). Koefisien yang nilainya negatif atau lebih dari 1 menunjukkan penyimpangan antara tujuan yang akan dites dan skor yang ditunjukkan peserta tes (Ebel, 1979).

### **Validitas Logis**

Validitas logis merupakan pembuktian kesesuaian isi pertanyaan dengan apa yang hendak diukur berdasarkan pengertian-pengertian logika. Karena itu, validitas ini tidak memerlukan alat uji berupa tes untuk menentukan kriterianya. Yang termasuk validitas logis ialah (1) validitas isi (*content validity*), (2) validitas konstruk (*construct validity*), dan (3) validitas penampakan (*face validity*). Yang dimaksud validitas isi ialah validitas suatu bentuk tes dilihat dari isi atau materi tes tersebut. Suatu tes memiliki validitas isi apabila materi/ isi tes tersebut merupakan bahan-bahan yang mewakili materi yang sudah diajarkan. Validitas isi bisa dianalisis menggunakan kisi-kisi soal.

Sebuah tes dikatakan memiliki validitas konstruk, menurut Heaton (1984:154) apabila tes tersebut dapat diukur ciri-ciri khusus tertentu

yang sesuai dengan teori tingkah laku bahasa dan teori belajar bahasa. Misalnya ada asumsi yang mengatakan bahwa kebiasaan berbahasa dapat diperoleh dari belajar grammar, maka tes yang memiliki validitas konstruk hendaknya disusun atas dasar penguasaan grammar.

Yang dimaksud validitas penampakan ialah ciri tes yang menunjukkan apakah suatu tes nampaknya mengukur apa yang hendak diukur (Djiwandono, 1996:97). Dalam penerapannya, ciri itu berupa hal-hal yang secara tampak luar memberikan kesan bahwa suatu tes memang dapat melakukan pengukuran yang seharusnya diukur. Misalnya, tes berbicara dikatakan memiliki validitas penampakan apabila peserta tes nampak saling berbicara dan bukan menulis.

## **2. Reliabilitas**

Reliabilitas diartikan sebagai keajegan atau kekonsistensian skor tes apabila tes tersebut diberikan kepada peserta lain yang memiliki tingkat kemampuan yang (hampir) sama (Gronlund, 1979; Oller, 1979; Weir, 1983; Djiwandono, 1996). Sebuah tes dikatakan tidak reliabel apabila setelah diberikan kepada peserta tes lain yang memiliki kemampuan hampir sama pada kesempatan yang berbeda, menghasilkan skor yang jauh berbeda (Brown, 1987).

Berbeda dengan validitas yang bisa diuji melalui analisis isi, reliabilitas menuntut piranti statistik untuk membuktikannya. Teknik pengujian reliabilitas ini mencakup dua hal: reliabilitas untuk soal yang memiliki skor dikotomus dan soal yang diskor secara berskala. Soal-soal yang dikotomus (salah = 0, benar = 1) diuji reliabilitasnya menggunakan metode tes ulang, metode tes setara, metode belah dua, metode Kuder-Richardson. Sedangkan soal essay yang diskor secara berskala diuji reliabilitasnya menggunakan metode koefisien alfa, dan metode antarpemilai (Djiwandono, 1996; Oller, 1979; Gronlund, 1979).

Tingkat koefisien reliabilitas juga bergerak dari 0-1. Semakin mendekati angka 1 semakin reliabel. Namun reliabilitas 1 dalam praktik hampir mustahil terjadi (Gronlund, 1979; Ebel, 1979; Djiwandono, 1996).

Koefisien reliabilitas dianggap sudah sangat baik apabila mencapai angka .90 ke atas. Jika terdapat koefisien negatif atau lebih dari satu, baik untuk validitas maupun reliabilitas, maka pasti terjadi proses yang salah.

Dengan kata lain, reliabilitas diartikan sebagai ciri tes yang memiliki kemampuan untuk menghasilkan pengukuran yang ajeg, tidak berubah-ubah seandainya digunakan secara berulang-ulang pada sasaran yang sama (Djiwandono, 1996:98). Karena itu, reliabilitas bukan terkait dengan tesnya melainkan terkait dengan skor yang ajeg yang diperoleh peserta tes. Artinya, skor yang diperoleh siswa dari tes tersebut stabil atau konsisten meskipun suasana pengujian atau penskroan yang dilakukan berbeda. Brown (1980:23) menyebutkan bahwa tes dikatakan tidak reliabel jika tes tersebut diberikan kepada peserta tes yang sama dalam waktu yang hampir sama, menghasilkan nilai yang jauh berbeda.

Berbeda dengan validitas yang pembuktiannya bisa dilakukan menggunakan logika, reliabilitas menuntut pembuktian menggunakan alat-alat statistik, misalnya menggunakan koefisien korelasi atau simpangan baku pengukuran. Djiwandono (1996:99) menyebutkan bahwa nilai koefisien reliabilitas bergerak dari 0 hingga 1. Semakin dekat dengan angka 1, semakin reliabel. Sedangkan koefisien dalam bentuk negatif atau lebih dari 1 menunjukkan reliabilitas yang menyimpang. Dalam praktik, angka reliabilitas tidak pernah mencapai angka 1. Untuk tes objektif pengujian reliabilitas di antaranya bisa dilakukan menggunakan: (1) metode tes ulang, (2) metode tes kembar, (3) metode behh dua, (4) rumus Kuder-Richardson (KR). Sedangkan untuk tes essay metode yang digunakan ialah (1) koefisien alfa, (2) koefisien antar penilai, dan (3) tabel kisi-kisi.

Metode tes ulang mempersyaratkan penggunaan tes yang sama dua kali, pada peserta tes yang sama. Metode tes setara memerlukan dua buah tes yang setara, yang mampu menghasilkan skor yang sama, tanpa mengulang penggunaan tes yang sama. Metode belah dua diperoleh dengan cara memberi tes sekali kepada peserta tes, kemudian jawaban peserta tes dibagi menjadi dua bagian, misalnya nomor ganjil dan nomor genap. Untuk jenis reliabilitas

ini, digunakan metode Spearman-Brown. Untuk tes yang diselenggarakan sekali saja, bisa dianalisis reliabilitasnya menggunakan metode KR, khususnya KR-20 atau KR-21. Penggunaan metode ini mempersyaratkan penggunaan skor diskrit, yaitu skor 1 untuk soal benar, dan skor 0 untuk soal salah.

Soal-soal essay bisa dicari reliabilitasnya menggunakan metode koefisien alfa, metode antar penilai, dan metode perkiraan. Soal yang diskor dengan cara skala memberikan nilai yang tidak sama untuk setiap butir, tergantung pada kesempumaan jawaban dan kelengkapannya. Untuk itu, metode koefisien alfa merupakan alat analisis yang cocok. Selain itu, untuk menghindari subjektivitas bisa juga digunakan metode antar penilai. Dalam hal ini nilai yang diberikan oleh dua orang penilai dikorelasikan.

Apabila cara statistik seperti tersebut di atas tidak bisa dilakukan, terutama untuk soal-soal buatan guru, metode analisis reliabilitas bisa dilakukan dengan cara perkiraan. Untuk ini, soal-soal dikelompokkan dalam katagori soal sulit, sedang, dan mudah. Katagori tersebut kemudian dibuat dalam tabel dan jumlah soal diproporsi sesuai dengan rencana penyusun soal.

Interpretasi koefisien validitas dan reliabilitas secara konvensional bisa dilakukan mengikuti tabel yang dikemukakan Djwandono (1996:154) berikut ini:

0.90 - 1,00	amat tinggi
0.70 - 0.89	tinggi
0.50 - 0.69	sedang
0.30 - 0.49	rendah
> 0.30	= amat rendah

### 3. Analisis Hasil Tes

Analisis hasil tes dimaksudkan untuk mengetahui informasi mengenai kemampuan siswa tertentu dalam menjawab tes. Informasi ini berupa skor yang sudah diinterpretasi menggunakan kriteria tertentu.

Untuk hasil yang memuaskan, analisis hendaknya disertai juga dengan informasi mengenai tingkat kesulitan ( $p$ ) setiap butir soal, daya pembeda ( $D$ ), validitas, dan reliabilitas. Untuk soal yang berbentuk pilihan ganda, analisis masih bisa ditambahkan dalam bentuk analisis efektivitas distraktor (pilihan jawaban).

### **Tingkat Kesulitan ( $p$ )**

Tingkat kesulitan tes menunjukkan seberapa sulit atau seberapa mudah butir-butir tes atau tes secara keseluruhan diselenggarakan (Djiwandono, 1996:140). Dengan demikian, akan diketahui apakah suatu butir tes tergolong sangat sulit, sulit, sedang, mudah, atau sangat mudah. Secara umum, tingkat kesulitan bisa dilihat dari skor rata-rata yang diperoleh siswa. Tingkat kesulitan bisa dinyatakan dalam bentuk persentase. Misalnya, tes yang dijawab benar oleh semua peserta tes, dinyatakan memiliki  $p = 100\%$  atau 1,00. Cara ini diperoleh dengan menggunakan rumus sederhana berikut:

$$p = \text{JJB} : \text{JPT} \times 100\%$$

dengan  $p$  = tingkat kesulitan, JJB = jumlah jawaban benar, sedangkan  $JPT$  = jumlah peserta tes.

Butir soal yang memiliki  $p$  mendekati 1 atau 100% berarti mudah atau sangat mudah. Sebaliknya, butir soal yang nilai  $p$ -nya mendekati 0 berarti sangat sulit atau sulit. Butir soal yang dianggap memiliki  $p$  ideal ialah 0.50, sedangkan  $p$  yang dianggap bisa diterima berkisar antara 0.20 - 0.80 (Djiwandono, 1996:141).

### **Daya Pembeda ( $D$ )**

Lebih lanjut, (Djiwandono, 1996:143-144) menjelaskan bahwa daya pembeda atau tingkat diskriminasi (disingkat  $D$ ) menunjukkan kemampuan tes dalam membedakan kemampuan antara peserta tes yang berkemampuan tinggi dan yang berkemampuan rendah. Semakin

tinggi D, semakin tinggi pula kemampuannya untuk membedakan kelompok siswa yang pandai dan yang kurang pandai. Tingkat diskriminasi 0 artinya soal tersebut tidak bisa membedakan siswa yang pandai dan yang kurang pandai. Tingkat diskriminasi  $D = 1,00$  berarti bahwa semua anggota kelompok pandai bisa menjawab benar, sementara anggota kelompok bawah tidak ada yang bisa menjawab soal dengan benar. Tingkat diskriminasi dilakukan dengan tahap-tahap berikut

1. Peketjaan peserta yang sudah dikoreksi dibagi ke dalam tiga kelompok atas dasar besarnya skor mentah. Ketiga kelompok itu ialah kelompok nilai tinggi, kelompok nilai sedang, dan kelompok nilai rendah. Penentuan jumlah kelompok atas dan bawah bisa berkisar antara 25% - 33% tergantung pada besarnya jumlah sampel.
2. Melakukan penghitungan jumlah jawaban benar terhadap masing-masing butir soal pada setiap kelompok, menggunakan rumus berikut

$$D = (T - R) : N$$

dengan  $D$  = nilai diskriminasi,  $T$  = jawaban benar pada kelompok tinggi,  $R$  = jawaban benar pada kelompok rendah, dan  $N$  merupakan jumlah peserta tes.

Dengan asumsi bahwa  $T$  selalu lebih pandai daripada  $R$ , maka  $D$  negatif diinterpretasikan sebagai indeks yang menyimpang, yang mungkin disebabkan oleh susunan atau rumusan butir tes yang tidak wajar. Penetapan indeks diskriminasi dilakukan menurut katagorisasi berikut ini:

0.50 atau lebih = baik

0.20 - 0.50 = sedang

> 0.20 = rendah

0 = tidak ada diskriminasi

negatif = menyimpang

## **BAB 4**

### **PENGEMBANGAN TES URAIAN**

#### **LATAR BELAKANG**

Tes uraian atau tes esai ialah tes yang pola jawabannya memerlukan penalaran dan logika atau argumen. Dalam memberikan argumen tersebut, penjawab memberikan dukungan data dan fakta yang dianalisis, disintesis, dan dievaluasi kecocokan dan kekuatannya dengan persoalan yang ditanyakan dalam butir tes. Tes uraian ini dalam perkembangan model pembelajaran dewasa ini, sejalan dengan model *authentic assessment* atau penilaian berbasis kelas.

Penggunaan tes uraian memiliki keuntungan dari dua aspek. *Pertama*, daya nalar yang dituntut oleh penjawab mengacu pada ranah berpikir yang lebih tinggi, mulai dari analisis, sintesis, dan analisis. Dengan ranah berpikir tinggi seperti ini, penjawab tes dituntut dan dilatih untuk berpikir kritis melalui analisis fakta dan data. Dengan demikian, jika penjawab memang tidak tahu tentang isi pertanyaan dalam soal, dia tidak akan bisa menjawab soal. *Kedua*, dari segi pembuatan butir pertanyaan, tes uraian relatif lebih mudah dirumuskan. Namun, tes uraian juga memiliki kelemahan. *Pertama*, tes uraian sulit dikoreksi karena memerlukan penguasaan materi yang diteskan, sistem skoring, dan kriteria penilaiannya. *Kedua*, tes uraian hanya bisa menjaring pertanyaan dalam jumlah sedikit, sehingga butir pertanyaan yang dirumuskan haruslah yang benar-benar penting.

Tes uraian, dilihat dari pola jawabannya, dikelompokkan menjadi tiga bagian. *Pertama*, tes uraian yang memerlukan jawaban relatif singkat dengan logika yang terikat pada konteks. Tes ini termasuk dalam katagori *close-ended test*. Tes model ini menuntut pemahaman yang kritis terhadap teks atau kasus yang digunakan sebagai latar dalam tes. *Kedua*, tes uraian yang memerlukan jawaban panjang yang harus disertai dengan fakta dan data. Tes ini termasuk dalam katagori *open-ended test*. Dalam memberi jawaban tes model ini, penjawab dituntut menunjukkan logika yang terarah, terbimbing oleh rujukan yang mapan, dan analisis kritis hasil inferensi beberapa *fakta dan rujukan yang dirangkum menjadi argumen yang mantap*. *Ketiga*, tes uraian bebas atau *extended writing*.

Tes ini berbentuk paper, telaah pustaka, ringkasan buku, artikel atau laporan hasil penelitian yang disiapkan dalam beberapa waktu. Jenis tes *close-ended*, cocok dikerjakan di dalam kelas dalam waktu antara 2-4 jam baik yang *open book* atau *close book*. Jenis tes *open-ended* dan *extended writing*, cocok digunakan untuk *take home test* yang memerlukan waktu antara 1-7 hari atau 3-8 minggu. Dalam mengerjakan jenis tes *close-ended*, umumnya penjawab tidak diperbolehkan membuka referensi apabila bahan tesnya sudah disiapkan sedemikian rupa. Tetapi pada kasus tertentu, tes *close-ended* ini bisa juga dijawab dengan membuka referensi saat mengerjakan tes. Untuk tes *open-ended* dan *extended writing*, penjawab justru Sangat dianjurkan untuk membaca referensi, mengikuti model yang sudah ditentukan, dan berkonsultasi terlebih dulu.

#### **TAHAPAN PENYUSUNAN BUTIR SOAL**

Dalam menyiapkan butir soal untuk tes uraian, secara umum diperlukan tiga langkah. Pertama, menyiapkan permasalahan yang akan dijadikan pertanyaan. Dalam menyusun permasalahan ini, sekaligus ditentukan indikator yang akan dijadikan *task*. Kedua, menyiapkan kunci jawaban. Kunci jawaban berisi pokok-pokok pikiran yang harus disampaikan sesuai dengan permasalahan. Kunci jawaban yang sebenarnya merupakan rambu-rambu jawaban ini memberi petunjuk kemendalaman dan kelengkapan jawaban. Ketiga, menyusun kriteria penilaian. Kriteria penilaian terutama diperlukan untuk jenis tes *extended writing*. Kriteria penilaian ini merupakan acuan yang dirujuk untuk menetapkan besaran skor atau hasil final hasil tes.

#### **CONTOH BUTIR PERTANYAAN**

Butir pertanyaan tes uraian, bisa dikembangkan dengan tiga model: (1) skor ditunjukkan langsung bersamaan dengan pertanyaan, (2) skor tidak ditunjukkan, dan (3) kriteria yang dievaluasi dijelaskan langsung. Untuk jenis tes *close-ended*, pola yang lebih mudah dan lazim digunakan ialah skor ditunjukkan bersamaan dalam butir pertanyaan.

(1) Skor 30

Dalam teori belajar dikenal pendekatan behavioristik, nativisme, dan interaksionisme. Jelaskan menggunakan contoh apa maksud dari setiap pendekatan tersebut. Jawaban harus mengacu pada landasan psikologi yang jelas dan akurat.

Kunci:

1. Definisi behavioristik + 1 contoh = 10;
2. Definisi nativisme + 1 contoh = 10;
3. Definisi interaksionisme + 1 contoh = 10;

- (2) Permasalahan pendidikan di Indonesia selalu terkait dengan kebijakan makro atau permasalahan yang muncul karena ketimpangan policy pemerintah dan permasalahan mikro atau permasalahan yang terjadi karena implementasi di lapangan. Pemerintah dipandang tidak tegas dalam membuat kebijakan pendidikan sehingga pelaksana pendidikan di tingkat provinsi dan kabupaten hanya bisa mengikuti saja kebijakan pemerintah pusat.

Tegaskan posisi Anda dalam menanggapi permasalahan ini. Jelaskan alasan Anda jika Anda setuju dan alasan Anda jika Anda tidak setuju. Jawaban Anda harus sejauh mungkin merujuk pada sumber-sumber referensi dan data yang akurat.

- (3) Buatlah sebuah artikel tentang Pembelajaran Berbasis Konteks berdasarkan psikologi konstruktivisme. Artikel harus ditulis mengikuti kriteria berikut ini.

- |                        |   |
|------------------------|---|
| Topik                  | : Pembelajaran Berbasis Konteks   |
| Panjang tulisan        | : 8-10 halaman spasi 2 ukuran kertas quarto (A-4)   |
| Sistematika artikel    | : a. Judul diikuti nama, NPM, kelas dan semester<br>b. Abstrak<br>c. Pendahuluan<br>d. Pembahasan<br>e. Penutup<br>f. Daftar Pustaka minimum 6 sumber |
| Aspek yang dievaluasi: | a. Organisasi tulisan<br>b. Substansi tulisan<br>c. Penggunaan kalimat dan peristilahan<br>d. Orisinalitas ide<br>e. Pengetikan atau mekanik tulisan  |

**CONTOH TEKS SOAL****UJIAN AKHIR**

**MATA KULIAH : BELAJAR DAN PEMBELAJARAN**  
**SEMESTER : 4**  
**PENGAJAR : Dr. Imroatus Solikhah, M.Pd**

**PETUNJUK**

Kerjakan dua soal di bawah ini seluruhnya. Jawaban Saudara harus mendalam, rinci, dan menunjukkan pemahaman yang lengkap mengenai suatu konsep. Beri ulasan dan alasan yang jelas untuk setiap uraian Saudara. Semakin mendalam tetapi jelas dan tidak bertele-tele, semakin bagus. Saudara diharapkan secara kritis menggunakan rujukan, referensi, dan kutipan-kutipan penting untuk mendukung jawaban Saudara. Jawaban harus diketik dalam spasi ganda dan dicetak di atas kertas quarto. Halaman sampul tidak diperlukan. Cantumkan identitas lengkap di bagian atas. Contoh,

**JAWABAN UJIAN AKHIR**  
**BELAJAR DAN PEMBELAJARAN**  
**TIARA LESTARI**  
**NIM. 200400000000**  
**Semester 4**

1. Pendekatan Pembelajaran Berbasis Konteks (*contextual teaching and learning*) di antaranya menganut aliran psikologi konstruktivisme dan menggunakan evaluasi proses yang autentik (*authentic assessment*).
  - a. Jelaskan secara mendalam menggunakan contoh apa yang dimaksud psikologi belajar konstruktivisme?
  - b. Jelaskan secara rinci perbedaan psikologi behaviorisme dan psikologi konstruktivisme dalam pembelajaran berbasis konteks.
  - c. Jelaskan secara rinci dan mendalam apa yang dimaksud *authentic assessment* dalam pembelajaran berbasis konteks? Apa saja jenis dan contoh *authentic assessment*?
  - d. Jelaskan dua alasan penting mengapa pembelajaran berbasis konteks harus menggunakan *authentic assessment*? Apa yang menyebabkan evaluasi produk tidak cocok dengan pembelajaran berbasis konteks?
  
2. Munculnya pendekatan pengajaran *integrative*, *communicative*, dan *contextual* terjadi di antaranya karena adanya anggapan bahasa tidak lagi dipandang sebagai sesuatu yang bersifat *discrete* dan bisa dipecah-pecah menjadi beberapa bagian. Ketiga metode mengajar tersebut saling menyempurnakan.
  - a. Jelaskan secara rinci menggunakan contoh apa perbedaan antara pendekatan mengajar *integrative*, *communicative*, dan *contextual*?

- b. Jelaskan apa yang dimaksud bahasa merupakan sesuatu yang diskrit dan dan apa dampaknya terhadap metode mengajar di kelas yang dipilih guru?
- c. Jelaskan minimal tiga hal yang menyebabkan metode *contextual* dianggap sebagai penyempurnaan metode *communicative*.

Bogor, 28 Maret 2006

## OPEN BOOK FINAL EXAM

<b>Course</b>	<b>: Thesis Writing</b>
<b>Semester</b>	<b>: VII</b>
<b>Duration</b>	<b>: 90 minutes</b>
<b>Instructor</b>	<b>: Prof. Dr. Teguh Budiharso, M.Pd</b>

### **PART A: Synthesizing Definitions of a Concept**

1. Before a problem statement or research question is determined, in the background of the study, the researcher should describe accurately (1) the history of the problems, (2) empirical evidence, and (3) theoretical evidence.

Explain the most essential definition on the history of problem, empirical evidence, and theoretical evidence. Your answer should be brief and concise.

2. The discussion section on research report normally presented in Chapter IV should cover: (1) explicit answer of the research problem and (2) confirmation of the findings of the recent study with previous studies stated in the review of literature.

Argue in brief and concise why explicit answer of research problem must be stated and why confirmation with literature review is required in a research report?

### **PART B: Developing Outline of Theoretical Review in Chapter II**

**Title 1:** A Study on the Dialects of Kutainese and Banjarnese Used for Verbal Daily

Communication among Staffs in Government Offices in Samarinda and Kutai Kartanegara.

**Title 2:** Teachers' Perspectives and Problems in Teaching English Using Competence

Based Curriculum (KBK) and Curriculum in the Lesson Unit Level (KTSP).

### **INSTRUCTION**

1. Pay considerable attention to the above two research titles. Write a plan for the review of related literature as if you wrote for Chapter II in the thesis report.
2. The plan should consist of list of topics only.

3. The structure of theoretical review should use approaches of Ontology, Epistemology, and Axiology.
4. The ontology refers to topics for sub-heading of the **Grand Theories**; the epistemology indicates topics of sub-headings of **Moderate Theories**; and the axiology should describe topics in the subheading of **Applied Theories**.
5. Each category in the grand theory, moderate theory, and applied theory may consist of one or more topics and sub-headings.

### **PART C: Composing an Essay, Writing Citation and References**

Select one of the two topics available below. Based on the topic develop an essay consisting of at least **five paragraphs**. Your essay should indicate high performance on the **maturity of composing quotations as for Chapter II** in a thesis. Indicate clearly how to write **good and correct citation in each paragraph and good reporting words**. After the end of your composition you should write **good and correct references**. Your essay must be based on the **three sources available** in this text.

#### **TOPICS**

1. Research problem and hypothesis in a research
2. Characteristics of good research problem and good hypothesis

#### **SOURCE 1: Problems and Problems Statement**

**Fred N. Kerlinger. 1973. Foundation of Behavioral Research. Quezon City, Philippines: Phoenix Press. Inc. page: 18.**

There are three criteria of good problems and problem statements. One, the problem should express a relation between two or more variables. For example, Is A related to B?. Two, the problem should be stated clearly and unambiguously in question form. The third criterion is the problem should imply possibilities of empirical testing. A problem that does not contain implications for testing its relation is not a scientific problem.

#### **SOURCE 2: Hypothesis**

**John W. Best. 1981. Research in Education. Englewood Cliffs, New Jersey: Prentice Hall Inc. page 6-7.**

After a problem has been identified, certain answers or solutions in the form of hypotheses may be formulated. The answers are based upon educated guess, past experience, informal observation. The hypothesis is a formal affirmative statement predicting a single research outcome, a tentative explanation of the relationship between two or more variables. It limits the focus of the investigation to a definite target and determines what observations are to be made.

**SOURCE 3: Definition of Problems**

**Donald Ary, Lucy Cheser Jacobs, Asghar Razavieh. 1985. Introduction to Research in Education. New York: The Dryden Press. Page: 9.**

A scientific inquiry originates from a problem or a question in need of solution. In scientific investigation, a question must have one essential characteristic: It must be possible to formulate it a way that observation or experimentation can provide an answer. The answer should be observed empirically by comparing with criteria or scores.

Puri Artha Hotel, Yogyakarta, 22 December 2006.

## BAB 5

### CONTOH PENGEMBANGAN TES MEMBACA

#### LATAR BELAKANG

Pengembangan tes termasuk keterampilan di bidang pengukuran dan evaluasi pembelajaran yang memerlukan ketekunan. Pertama, pengembangan tes yang melibatkan analisis validitas item tidak mutlak diperlukan untuk pengukuran hasil belajar harian (formatif tes). Dalam proses pembelajaran guru dianggap sudah menguasai peta pemahaman pembelajar sehingga bentuk tes yang didesain sebagai *classroom test* dianggap sudah memadai (Brown, 1997; Djiwandono, 1997). Kedua, pembuatan tes yang melibatkan uji-coba untuk analisis item dan validitas butir soal memerlukan waktu yang cukup panjang dan keterampilan statistik yang relatif rumit. Karena itu, tidak semua pengajar memiliki kesempatan melakukan uji coba butir soal dan bahkan sebagian lainnya tidak telaten karena kurang menguasai teknik analisisnya (Budiharso, 2005).

Untuk mengatasi persoalan tersebut, pembelajar dan praktisi pendidikan biasanya menggunakan strategi khusus yang dianggap mudah walaupun strategi tersebut secara akademik belum bisa dipercaya sepenuhnya. Brown (1997) menggunakan istilah validitas prediksi, Djiwandono (1995) menggunakan istilah validitas pengalaman untuk menetapkan tingkat kesulitan butir. Validitas ini digunakan untuk menetapkan butir soal yang dianggap sesuai dengan tujuan tes dengan cara mengidentifikasi tingkat kesulitan ke dalam bentuk mudah, sedang, dan sulit. Penentuan tingkat kesulitan tersebut didasarkan pada penggunaan kata kerja operasional yang diusulkan dalam taksonomi Bloom untuk menunjukkan tingkatan berpikir dalam ranah kognitif. Bloom membagi ranah kognitif ke dalam enam tingkatan: *recall*, *comprehension*, *application*, *analysis*, *synthesis*, dan *evaluation*. Setiap ranah diukur menggunakan kata kerja operasional seperti, mengidentifikasi untuk tingkat *recall*, menjelaskan untuk tingkat *application*, membedakan untuk tingkat evaluasi, dan seterusnya.

Cara lain yang juga lazim digunakan dalam menentukan tingkat kesulitan butir berdasarkan pengalaman ialah dengan membuat kisi-kisi butir soal (Grondlund, 1997; Djiwandono, 1995; Budiharso, 2005). Untuk tujuan tertentu, praktisi pendidikan yang menyelenggarakan tes secara masal seperti ujian nasional, penentuan butir soal diidentifikasi melalui teknik tes daya serap. Tes daya serap dilakukan dengan cara mengundang peserta tes, misalnya murid SMP atau SMA di berbagai wilayah. Jawaban siswa digunakan untuk menggambarkan seberapa banyak siswa yang dianggap mampu dan diprediksi akan lulus dalam ujian nasional yang sebenarnya dan untuk menetapkan jenis butir soal mana yang mudah dijawab karena banyak siswa yang menjawab benar atau sebaliknya butir soal yang sulit dilihat dari sedikit siswa yang bisa menjawab benar (Budiharso, 2007).

Cara yang diuraikan di atas dalam perspektif pengujian dan evaluasi kurang dapat dipertanggungjawabkan. *Pertama*, tidak ada bukti empiris bahwa butir-butir soal yang disusun sesuai dengan tingkat kesulitan yang sesuai. *Kedua*, penentuan proporsi butir soal ke dalam katagori mudah, sedang, dan sulit baru bisa dianggap akurat apabila butir-butir soal tersebut diuji-cobakan lebih dulu. Karena itu, teknik penentuan kualitas butir soal yang dilakukan melalui ujicoba tetap harus dilakukan.

Dalam penelitian, ada proses pengembangan instrumen yang di dalamnya mengharuskan peneliti untuk melakukan ujicoba. Langkah-langkah pengembangan instrumen penelitian tersebut sudah baku dan secara ketat harus dilakukan peneliti. Tujuan pelaksanaan uji coba tersebut ialah terutama variabel yang diukur melalui indikator yang harus diidentifikasi dan dijelaskan secara tepat oleh peneliti. Melalui pengembangan instrumen, peneliti juga dituntut merumuskan kerangka pikir penelitian, definisi konsep dan definisi operasional yang akan menuntun peneliti dalam mengambil data. Karena itu, esensi pelaksanaan uji-coba butir soal sangat penting dalam penelitian dan peneliti dituntut untuk melakukan rangkaian uji coba instrumen melalui tahapan-tahapan yang terstandar.

Makalah ini mendeskripsikan proses pengembangan butir soal membaca pemahaman untuk SMA. Rangkaian uji coba yang dilakukan meliputi tahap

pengembangan butir, uji-coba pertama untuk melihat tingkat kesulitan, dan ujicoba kedua untuk melihat validitas dan reliabilitas butir. Membaca pemahaman dipilih sebagai fokus karena pertimbangan praktis saja. Selain teks mudah dikembangkan dalam bentuk butir soal, peneliti memiliki pengalaman empiris dalam hal mengajar membaca pemahaman. Kedua alasan tersebut dianggap memudahkan peneliti untuk memudahkan menyusun butir soal dan menganalisis hasil ujicoba.

### **PERMASALAHAN**

Permasalahan yang dikembangkan dalam laporan ini ialah analisis butir soal tes pengembangan. Secara umum, analisis diarahkan untuk melihat validitas butir dan reliabilitas total seluruh butir soal. Secara spesifik, analisis diarahkan untuk melihat lima hal: tingkat kesulitan butir, daya pembeda, efektivitas distraktor, validitas total, dan reliabilitas total.

### **LANDASAN TEORI MEMBACA PEMAHAMAN**

Membaca hakikatnya ialah proses membangun pemahaman wacana tulis. Proses ini terjadi dengan menjodohkan atau menghubungkan skemata pengetahuan dan pengalaman yang telah dimiliki sebelumnya sehingga membentuk pemahaman terhadap wacana yang dibaca (Syafiie, 1999:35). Pemahaman dalam membaca dikelompokkan ke dalam kelompok membaca rendah dan kelompok membaca tinggi. Kelompok membaca rendah disebut juga membaca literal. Kelompok membaca tinggi disebut juga membaca kritis, dan terdiri dari tiga: pemahaman interpretif, pemahaman kritis, dan pemahaman kreatif (Burns, Roe, dan Ross, 1996). Tindakan membaca terdiri dari dua bagian: proses membaca dan hasil membaca. Sebagai proses, membaca melibatkan juga aktivitas mental dan fisik. Sedangkan hasil membaca mengacu pada komunikasi pikiran atau perasaan penulis kepada pembaca.

Dalam proses membaca, terdapat empat level pemahaman, yaitu: (1) pemahaman literal, (2) pemahaman interpretif, (3) pemahaman kritis, dan (4) pemahaman kreatif (Burns, 1980:369). Setiap level ini dianggap sebagai jenis

kemampuan tersendiri dan dalam prosesnya untuk mencapai level yang lebih tinggi selalu melewati level di bawahnya (Nurhadi, 1987:72). Pengembangan butir tes ini dirancang untuk mengukur pemahaman membaca literal saja karena alasan praktis untuk kemudahan dan ketersediaan waktu yang terbatas. Analisis membaca literal dianggap memadai untuk tujuan praktik pengembangan instrumen dan analisis hasil uji coba butir soal.

Pemahaman yang paling rendah ialah membaca pemahaman literal. Pemahaman literal mengacu pada apa yang disebutkan dalam teks bacaan. Tujuan pemahaman literal ialah untuk memperoleh detil isi bacaan secara efektif.

Pemahaman interpretif dihasilkan melalui proses berpikir yang lebih tinggi, yaitu interpretasi, analisis, dan sintesis suatu bacaan. Pemahaman interpretif menuntut pembaca membuat interpretasi isi wacana walaupun informasi yang diperlukan tidak terdapat dalam teks. Pemahaman dibentuk dengan cara membuat analisis atau inferensi. Dalam pemahaman jenis ini prediksi isi bacaan ialah keterampilan yang penting, sebab dengan memprediksi seseorang dapat membuat hipotesis dilanjutkan dengan penerimaan, pengakuan, atau penolakan terhadap isi hipotesis tersebut.

Pemahaman kritis bertujuan untuk memberi penilaian terhadap isi teks yang dibaca. Penilaian dilakukan dengan cara membandingkan dengan pengetahuan yang sudah dimiliki atau sumber lain. Dalam membaca jenis ini, pembaca bisa memberi komentar, menemukan kekurangan, atau menjelaskan keunggulan isi teks.

Tingkat pemahaman kreatif menuntut pembaca untuk menemukan ide dalam teks yang dianggap kurang benar, kurang akurat, atau bertentangan dengan sumber lain. Jika dalam membaca kritis, pembaca cukup memberi komentar atau pendapat lain, dalam membaca kreatif, pembaca bisa membuat telaah, summary, atau kompilasi hasil bacaan.

### **Membaca Literal**

Pemahaman literal ialah pemahaman terhadap apa yang disebutkan dalam teks bacaan. Sebagai keterampilan paling dasar dalam pemahaman, pemahaman literal

bertujuan untuk memperoleh informasi langsung dalam teks bacaan. Syafi'ie (1993:48) menjelaskan pemahaman literal adalah keterampilan memahami isi bacaan sebagaimana yang tertulis melalui pemahaman arti kata, kalimat, dan paragraf dalam wacana. Pemahaman ini hanya menuntut kemampuan ingatan, yaitu ingatan mengenai apa yang tertulis dalam bacaan. Pemahaman literal dimaksudkan untuk memahami urutan waktu yang ditandai dengan kata sekarang, sebelum, ketika, sementara, masih, setelah atau hubungan sebab akibat yang ditandai dengan sebab, karena, dengan demikian. Menurut Dorothy (1995) pertanyaan literal meminta jawaban yang lebih sederhana, meminta pembaca mengingat kembali memori yang terdapat dalam teks.

Kriteria pemahaman literal yang terpenting ialah mengenal dan mengingat apa yang tertulis dalam bacaan. Untuk membangun pemahaman literal ini, pembaca dapat menggunakan arahan dengan menggunakan kata tanya: (1) siapa, untuk menanyakan siapa atau tokoh dalam wacana, (2) apa, untuk menanyakan barang, benda, atau peristiwa, (3) di mana, untuk menanyakan tempat, (4) kapan, untuk menanyakan waktu, (5) bagaimana, untuk menanyakan proses jalannya peristiwa, (6) mengapa, untuk menanyakan alasan sesuatu yang disebutkan dalam teks bacaan (Burns, Roe, dan Ross, 1996:255).

Kriteria lain dalam membaca pemahaman menurut Syafi'ie (1993:48) ialah menggunakan pertanyaan yang berhubungan dengan ingatan. Kata-kata tersebut antara lain: (1) pertanyaan tentang fakta dan detil bacaan, (2) peristiwa dan urutan kejadiannya, (3) hal-hal penting yang sering diulang dalam bacaan, (4) mengecek makna yang sesuai, (5) pertanyaan tentang ide pokok kalimat dalam paragraf.

### **Jenis Pertanyaan dalam Membaca Pemahaman**

Pertanyaan dalam membaca pemahaman yang digunakan dalam laporan ini mengacu pada rumusan yang dikembangkan oleh Shape (2000:316) dalam Baron's TOEFL. Shape (2000) mengidentifikasi delapan jenis pertanyaan yang lazim digunakan dalam tes membaca. Ke-8 jenis tersebut ialah: preview (*previewing*),

pertanyaan untuk mengetahui ide pokok (*reading for main ideas*), pertanyaan kosa kata dalam konteks (*using context for vocabulary*), skaning untuk mencari detail (*scanning for details*), membuat inferensi (*making inferences*), mengidentifikasi perkecualian (*identifying exceptions*), menemukan referens (*locating references*), dan pertanyaan yang merujuk kembali pada teks (*referring to the passage*).

Pertanyaan preview bertujuan untuk mengetahui ide umum suatu teks. Pertanyaan ini bertujuan untuk mengetahui ide umum yang tertuang dalam teks; untuk mengetahui kesan mengenai isi teks. Preview biasanya dikerjakan dengan cara membaca kalimat awal dan kalimat akhir pada setiap paragraf.

Membaca untuk mengetahui ide utama bertujuan untuk mengetahui pandangan atau maksud penulis. Isi pesan bisa ditangkap melalui pernyataan *thesis* dalam teks. Secara spesifik, ide utama bisa diketahui dari topik apa yang dibicarakan oleh penulis dalam teks. Jenis pertanyaan untuk mengetahui ide utama bisa berupa: (1) Apa ide utama teks?, (2) Apa subjek yang ditulis?, (3) Apa topik bacaan?, dan (4) Judul apa yang paling cocok?

Dalam pertanyaan makna kosa kata yang mengacu ada konteks, pertanyaan bisa diketahui jawabannya dengan melihat pada kata, frase, atau kalimat yang ada di sekitar kata yang ditanyakan maknanya. Itulah yang disebut konteks, yang berupa kata atau kalimat dalam bacaan. Konteks berfungsi untuk menebak makna. Umumnya, jika pembaca mengetahui makna umum suatu kalimat, dia juga akan tahu makna makna umum dalam kalimat. Jadi, menebak makna melalui konteks bisa digunakan untuk menebak makna kosa kata tertentu yang ditanyakan dalam soal.

Dalam pertanyaan skaning, pertanyaan diarahkan untuk mengetahui informasi detail tertentu. Untuk menjawab pertanyaan jenis ini, ada tiga langkah yang diperlukan. Pertama, pembaca harus menemukan *content words* berupa kata benda, kata kerja, kata sifat, dan kata keterangan yang dianggap penting. Kedua, pembaca harus mencari kata yang memiliki sinonim atau antonim dari kata yang dicari. Dengan cara membaca cepat, pembaca akan menemukan tempat kata yang

ditanyakan dalam teks. Selanjutnya, pembaca harus mengkaji kalimat yang ditanyakan sesuai dengan bunyi pertanyaannya.

Pertanyaan inferensi ialah pertanyaan yang keterangannya tidak dinyatakan langsung dalam teks. Jadi pembaca harus mengaitkan dengan keterangan dalam teks atau informasi di luar teks. Biasanya yang terdapat dalam teks ialah fakta atau pernyataan tertentu yang bisa dikaitkan satu sama lainnya. Jadi pertanyaan inferensi dibuat berdasarkan penarikan simpulan logis menggunakan fakta dalam teks atau pendapat penulis.

Pertanyaan yang mengacu pada perkecualian berisi beberapa pernyataan yang dianggap semua salah atau semua benar; dan dari pernyataan tersebut terdapat satu yang merupakan perkecualian yang merupakan jawaban yang diminta. Pertanyaan jenis ini bisa dijawab menggunakan teknik skaning.

Mengidentifikasi referens dimaksudkan untuk mengetahui pelaku atau kata ganti yang sesuai dengan konteks. Umumnya, pertanyaan dimulai dengan kata ganti: it, he, she, dsb, dan pembaca diminta untuk menetapkan kata ganti yang paling cocok dengan konteks. Teknik menjawab pertanyaan jenis ini meliputi menemukan lebih dulu kata ganti yang ditanyakan dan temukan kata ganti dalam ada dalam pertanyaan dan pilihan jawaban. Makna kalimat tidak akan berubah kalau pembaca menemukan kata ganti yang tepat.

Jenis pertanyaan terakhir ialah menetapkan isi teks. Pertanyaan jenis ini ditandai dengan cara menanyakan makna baris tertentu dalam teks. Biasanya, pertaynaannya digaris bawah. Teknik yang bisa digunakan untuk menjawab pertanyaan ialah pembaca harus mengidentifikasi lebih dulu bunyi pertanyaannya, lalu temukan makna kata atau kalimat yang digarisbawahi dalam teks.

## **KERANGKA PENGEMBANGAN**

Pengembangan butir soal ini mengikuti kerangka pikir pengembangan tes dari Djiwandono (1995), Brown (1997), Djaali (2008), dan Valencia, Hiebert, dan Afflerbach (1996). Pengembangan dimulai dari sintesis teori pendukung,

penyusunan definisi konsep, perumusan definisi operasional, pembuatan kisi-kisi, dan pengembangan butir soal. Setelah butir soal dikembangkan, butir soal diujicobakan beberapa kali. Ujicoba pertama, digunakan untuk mengidentifikasi tingkat kesulitan butir (*item degrees of difficulty*), daya pembeda (*discrimination*), efektivitas distraktor (untuk butir soal dengan bentuk pilihan ganda). Apabila butir soal yang dianggap sesuai dengan kriteria tingkat kesulitan yang diperlukan sudah diperoleh, tes tersebut diujicobakan lagi untuk mengetahui tingkat validitas dan reliabilitas. Kerangka pikir ini dikembangkan dan mudah diikuti untuk semua pihak terutama yang menggunakan analisis secara manual. Walaupun rumit, analisis ini memiliki keunggulan karena pelacakan butir bisa lebih teliti, mulai dari tingkat kesulitan, daya pembeda, dan distraktor.

Analisis model lain yang juga banyak dilakukan, misalnya dianjurkan oleh Arikunto (2007) dan Djaali (2008) karena dianggap lebih praktis ialah analisis validitas butir dan reliabilitas total. Untuk analisis menggunakan program tertentu misalnya SPSS, analisis ini banyak dianjurkan. Namun, analisis ini memiliki kelemahan karena tidak bisa melacak tingkat kesulitan, daya pembeda, tingkat efektivitas distraktor butir soal. Informasi yang dicakup dalam validitas butir dianggap mewakili keterangan tingkat kesulitan dan daya pembeda. Sedangkan analisis distraktor tidak dilakukan.

Dengan alasan demikian, analisis dalam pengembangan butir soal ini dilakukan melalui dua cara: teknik manual dan teknik program SPSS Release 11. Kedua model analisis ini dianggap memadai karena sekaligus bisa memberikan informasi lengkap yang meliputi: tingkat kesulitan butir, daya pembeda, efektivitas distraktor, validitas, dan reliabilitas.

### **Definisi Konseptual**

Membaca pemahaman literal ialah penguasaan isi bacaan yang diperoleh melalui pemahaman pertanyaan terbimbing seperti apa, siapa, kapan, di mana, dan

bagaimana yang bertujuan untuk mengingat dan menjelaskan kembali isi bacaan sesuai dengan yang tertuang dalam teks.

### **Definisi Operasional**

Membaca literal ialah identifikasi makna bacaan untuk mengingat isi teks melalui pertanyaan pokok pikiran, ide utama, makna kosa kata, makna konteks, kesesuaian referens, perkecualian pernyataan, dan makna kalimat dalam teks.

### **Kisi-Kisi Soal**

Kisi-kisi yang dimaksudkan di sini untuk menunjukkan 25 butir soal yang diperoleh setelah uji-coba. Soal 25 butir itu dikembangkan dari 6 teks. Sebelum uji coba, jumlah butir sebanyak 30 teks dikembangkan dari 7 teks. Kisi-kisi soal disajikan dalam Tabel 1.

### **Proses Ujicoba**

Ujicoba butir soal ini dilakukan sebanyak 3 kali. Ujicoba dilakukan masing-masing pada siswa kelas III-A, III-B, III-C MAN 1 Blitar. Setiap kelompok ujicoba digunakan 10 responden. MAN 1 Blitar dipilih karena pertimbangan kepraktisan dan akses. Peneliti memiliki akses ke sekolah tersebut sehingga memudahkan pelaksanaan ujicoba. Sedangkan responden 10 orang digunakan sesuai saran dari Brown (1995), Djiwandono (1995) dan Grounlund (1979). Tujuan ujicoba ini ialah untuk mengetahui kualitas butir soal, sehingga yang terpenting ialah tersedianya jumlah peserta tes yang memiliki kriteria sesuai dengan rancangan tes. Untuk kebutuhan tersebut, 10 orang sudah dianggap cukup (Brown, 1995; Djiwandono, 1995; Grounlund, 1979).

Tabel 1. Kisi-Kisi Soal Setelah Ujicoba

No	Teks	Nomor Soal	Jenis Pertanyaan	Jumlah Tiap Passage
1	Cassava	1	Preview	6
		2	Main idea	
		3	Scanning	
		4	Inferences	
		5	Exception	
		6	Referring passage	
2	Measles	7	Exception	2
		8	Scanning	
3	Forest	9	Preview	5
		10	Scanning	
		11	Exception	
		12	Inferences	
		13	Main idea	
4	The Digestive System	14	Vocabulary	3
		15	Scanning	
		16	Inferences	
5	Our Body	17	Referring passage	5
		18	Scanning	
		19	Scanning	
		20	Preview	
		21	Referring passage	
6	Badminton	22	Exception	4
		23	Exception	
		24	Inferences	
		25	Exception	
			Jumlah	25

Ujicoba pertama, dimaksudkan untuk menentukan kualitas butir yang meliputi: tingkat kesulitan butir, daya pembeda, dan afektivitas distraktor. Peserta yang digunakan sebagai responden ialah kelas III-A. Dari hasil ujicoba pertama, diperoleh butir soal sebanyak 25 butir; yang berarti 5 butir dari sebelumnya berjumlah 30 dibuang. Kelima butir yang dibuang tersebut ialah seluruh pertanyaan pada teks pertama.

Setelah hasil analisis kualitas butir diketahui, butir soal direkonstruksi lagi dan diperoleh 25 butir yang dikembangkan dari 6 teks. Butir soal ini dikemas lagi menjadi seperangkat tes dan diujicobakan lagi kepada 10 orang siswa Kelas III-B. Tujuan ujicoba yang kedua ini ialah untuk mengetahui validitas dan reliabilitas ke-25 butir soal tersebut (Brown, 1995; Djiwandono, 1995). Uji validitas dilakukan menggunakan teknik kriteria; nilai ujicoba dibandingkan dengan nilai guru yang diperoleh dari tes membaca pemahaman. Sedangkan uji reliabilitas dilakukan menggunakan teknik test-retest. Untuk memperoleh skor uji reliabilitas, tes yang sama diberikan kepada peserta yang sama dalam waktu yang hampir bersamaan (Djiwandono, 1995). Dengan demikian, siswa kelas III-B dites sebanyak 2 kali dalam waktu berselang 30 menit dari tes pertama dan tes kedua.

Karena analisis ini menggunakan dua teknik, manual dan SPSS Release 11, peneliti secara sengaja mengadakan tes yang keempat kalinya. Tes ini diberikan pada siswa kelas III-C. Butir tes yang digunakan ialah butir soal yang sudah dianalisis kualitas butirnya sebanyak 25 item. Kelas A, B, dan C sengaja dites terpisah untuk menghindari kebocoran pertanyaan dan menjaga validitas eksternal. Selain itu, analisis yang digunakan ialah manual dan SPSS, karena itu, diperlukan jenis data yang berbeda. Kelas A, B, dan C di MAN 1 Blitar dianggap sama kemampuannya berdasarkan informasi dari sekolah dan data akademik siswa.

### **Teknik Analisis**

Teknik analisis yang digunakan dalam laporan ini ada dua jenis: manual dan teknik SPSS Release 11.

#### ***Teknik Manual***

Teknik manual digunakan untuk mengetahui lima hal: tingkat kesulitan butir item, daya pembeda item, efektivitas distraktor, validitas, dan reliabilitas. Tingkat kesulitan butir item menunjukkan taraf kesulitan butir dalam tiga kategori: sedang,

sulit, dan sangat sulit. Menggunakan kriteria dari Djiwandono (1995) dan Brown (1995), kriteria tingkat kesulitan yang digunakan ialah 20% sd 80%. Semakin tinggi persentasenya menunjukkan butir mudah dan semakin rendah mendekati 20% menunjukkan butir soal sulit. Kriteria tingkat kesulitan yang diusulkan Djiwandono (1995) dan Brown (1995) ialah:

81-100%	Sangat mudah
70-80	Mudah
41-69	Sedang
20-40	Sulit
10-20	Sangat sulit

Daya pembeda menunjukkan kemampuan butir soal dalam membedakan peserta tes yang mengerti dan yang tidak mengerti. Mengikuti Dhiwandono (1995) dan Brown (1995), kriteria diskriminasi yang digunakan ialah 20% sd 100%. Kebalikan dari tingkat kesulitan, semakin mendekati 100%, daya pembeda butir semakin bagus. Brown (1995) dan Djiwandono (1995) mengusulkan kriteria diskriminasi berikut ini.

100%	Diskriminasi sempurna
50%	Ideal
20%	Sedang
0%	Tidak ada diskriminasi
(-) Negatif	Diskriminasi terbalik (kelompok rendah menjawab benar dan kelompok tinggi justru menjawab salah)

Analisis validitas yang digunakan ialah validitas kriteria. Validitas maksudnya kesesuaian atau kecocokan antara nilai tes dan kemampuan peserta tes (Djiwandono, 1995). Dalam laporan ini, analisis validitas kriteria memerlukan dua set skor yang diperoleh dari skor tes ujicoba dan skor pembandingan dari tes lain yang menguji hal yang sama. Dalam laporan ini, skor pembandingan diperoleh dari nilai siswa Kelas III-

B yang mengikuti ujicoba kedua. Skor diberikan oleh guru yang juga diperoleh dari nilai membaca pemahaman. Analisis digunakan rumus Product Moment.

Analisis reliabilitas yang digunakan dalam laporan ini ialah analisis menggunakan teknik test-retest. Reliabilitas maksudnya keajegan atau konsistensi dari skor pada tes pertama dibandingkan skor pada tes kedua. Skor tes pertama diperoleh dari tes pada ujicoba kedua yang diberikan pada kelas III-B. Sedangkan skor kedua diperoleh juga dari kelas III-B estela mereka diberi waktu 25 menit untuk istirahat dari tes pertama, untuk mengerjakan tes yang sama yang kedua kalinya. Sama dengan analisis validitas, analisis reliabilitas juga menggunakan humus product Moment. Kriteria penentuan validitas dan reliabilitas dinyatakan dalam bentuk persen untuk memudahkan interpretasi berapa persen tingkat akurasi. Kriteria yang dikemukakan Djiwandono (1995) dan Brown (1995) disajikan berikut ini.

80%-100%	Sangat tinggi
70%-79%	Tinggi
60%-69%	Sedang
40%-59%	Kurang
10%-39%	Sangat kurang

### ***Analisis Menggunakan SPSS Release 11***

Taknik analisis menggunakan program SPSS Relase 11 digunakan sebagai pembanding. Analisis ini bertujuan untuk mengetahui validitas butir dan reliabilitas total seluruh item. Program SPSS Release 11 dipilih karena program ini termasuk yang mutakhir dan peneliti menguasai program ini.

Analisis menggunakan SPSS Release 11 memerlukan cukup satu set skor. Skor tersebut diambil dari tes yang sudah diseleksi tingkat kesulitannya sebanyak 25 butir. Untuk membedakan dengan hasil analisis manual, peneliti sengaja menggunakan data yang berbeda yang diperoleh dari ujicoba pada kelas III-C. Hasil analisis SPSS memberikan informasi sejauh mana hasil uji korelasi setiap butir lebih besar atau lebih kecil dari kriteria. Jika hasil uji statistik suatu butir lebih kecil dari kriteria, butir tersebut dianggap tidak valid dan tidak bisa digunakan. Sebaliknya,

hasil uji statistik butir soal yang lebih tinggi dari kriteria, dianggap valid dan direkomendasikan untuk digunakan.

## **HASIL UJI COBA**

Hasil ujicoba laporan ini berisi produk butir soal sebanyak 25 dan hasil analisis statistik. Hasil berupa produk dalam bentuk butir tes disajikan dalam Lampiran 1, sedangkan hasil analisis diuraikan berikut ini.

### **Hasil Analisis Teknik Manual**

Analisis teknik manual pada tahap awal digunakan untuk menganalisis tingkat kesulitan butir soal, daya pembeda, dan efektivitas distraktor. Pada tahap berikutnya analisis digunakan untuk menguji validitas dan reliabilitas.

### ***Tingkat Kesulitan Butir Soal***

Tingkat kesulitan butir soal dimaksudkan untuk mengetahui bagaimana kualitas setiap item dalam katagori mudah, sedang, sulit. Seperti disebutkan sebelumnya, data untuk analisis ini diambil dari ujicoba tahap pertama yang diberikan kepada 10 orang dari klas III-A yang dipilih secara random. Data pada Tabel 2 menunjukkan hasil analisis butir soal. Hasil analisis menunjukkan sebanyak 10 (30%) dalam katagori mudah, 6 (24%) sedang, dan 9 (36%) sulit. Ini menunjukkan bahwa proporsi tingkat kesulitan butir terpenuhi.

Tabel 2. Analisis Tingkat Kesulitan Butir Soal

No	Tingkat Kesulitan	Katagori	Jumlah Butir	Persentase
1	20	Sulit	3	12%
2	30	Sulit	2	8%
3	40	Sulit	4	16%
4	50	Sedang	2	8%
5	60	Sedang	4	16%
6	70	Mudah	6	24%
7	80	Mudah	4	16%
			25	100%

Seperti nampak dalam Tabel 3, butir soal yang dianalisis sebanyak 30 item. Dari jumlah itu, terdapat 5 butir soal (1 sd 5) yang tidak memenuhi kriteria sehingga tidak digunakan, sedangkan item nomor 6 sd 25 masuk kriteria yang bisa digunakan.

Tabel 3. Hasil Analisis Tingkat Kesulitan Butir Soal

No of items	Correct	Wrong	p	Criteria
1	0	10	0.00	<b>Too hard</b>
2	10	0	1.00	<b>Too easy</b>
3	9	1	0.90	<b>Too easy</b>
4	10	0	1.00	<b>Too easy</b>
5	10	0	1.00	<b>Too easy</b>
6	4	6	0.60	<b>Used</b>
7	3	7	0.70	<b>Used</b>
8	3	7	0.30	<b>Used</b>
9	2	8	0.20	<b>Used</b>
10	2	8	0.20	<b>Used</b>
11	6	4	0.60	<b>Used</b>
12	5	5	0.50	<b>Used</b>
13	4	6	0.40	<b>Used</b>

14	7	3	0.70	<b>Used</b>
15	2	8	0.80	<b>Used</b>
16	2	8	0.80	<b>Used</b>
17	4	6	0.40	<b>Used</b>
18	3	7	0.30	<b>Used</b>
19	6	4	0.60	<b>Used</b>
20	6	4	0.60	<b>Used</b>
21	7	3	0.70	<b>Used</b>
22	7	3	0.70	<b>Used</b>
23	5	5	0.50	<b>Used</b>
24	4	6	0.40	<b>Used</b>
25	2	8	0.20	<b>Used</b>
26	4	6	0.40	<b>Used</b>
27	3	7	0.70	<b>Used</b>
28	8	2	0.80	<b>Used</b>
29	7	3	0.70	<b>Used</b>
30	8	2	0.80	<b>Used</b>

### ***Hasil Analisis Daya Pembeda***

Analisis Daya Pembeda dimaksudkan untuk mengetahui efektivitas butir dalam membedakan kemampuan peserta tes yang menguasai materi dan yang tidak menguasai (*the better and poorer ones*). Hasil analisis pada Tabel 4 menunjukkan bahwa dari 30 butir yang dianalisis, 5 butir tidak memiliki diskriminasi yang diharapkan. Hasil ini sejalan dengan analisis kualitas butir yang menunjukkan bahwa item nomor 1 sd 5 tidak bisa digunakan.

Tabel 4. Hasil Analisis Diskriminasi

No Item	Jawaban Atas	Jawaban Bawah	Selisih	D	Kriteria
1	5	5	0	0	Tdk dipakai
2	4	6	-2	-0.4	Tdk dipakai
3	5	5	0	0	Tdk dipakai
4	3	5	-2	-0.4	Tdk dipakai
5	5	5	0	0	Tdk dipakai
6	4	1	3	0.6	Dipakai

Tabel 4. Hasil Analisis Diskriminasi

No Item	Jawaban Atas	Jawaban Bawah	Selisih	D	Kriteria
7	4	2	2	0.2	Dipakai
8	5	1	4	0.8	Dipakai
9	5	0	1	1.0	Dipakai
10	5	1	4	0.8	Dipakai
11	4	0	4	0.8	Dipakai
12	5	0	5	1.0	Dipakai
13	4	2	2	0.4	Dipakai
14	5	3	2	0.4	Dipakai
15	3	2	1	0.2	Dipakai
16	3	1	2	0.4	Dipakai
17	4	1	3	0.6	Dipakai
18	4	0	4	0.8	Dipakai
19	5	2	3	0.6	Dipakai
20	5	1	4	0.8	Dipakai
21	5	4	1	0.2	Dipakai
22	5	4	1	0.2	Dipakai
23	5	0	5	1.0	Dipakai
24	4	1	3	0.6	Dipakai
25	4	0	4	0.8	Dipakai
26	3	0	3	0.6	Dipakai
27	3	1	2	0.4	Dipakai
28	5	4	1	0.2	Dipakai
29	5	3	2	0.4	Dipakai
30	5	3	2	0.4	Dipakai

### *Analisis Tingkat Kesulitan dan Daya Beda*

Untuk lebih memastikan bahwa tiap butir memiliki kualitas yang baik dari segi tingkat kesulitan dan daya beda, disajikan juga analisis gabungan keduanya. Data hasil analisis disajikan dalam Tabel 5 berikut.

Tabel 5. Analisis p dan D

No Item	p (%)	D (%)	Keterangan
1	0	0	Tidak dipakai
2	100	-40	Tidak dipakai
3	90	0	Tidak dipakai
4	100	-40	Tidak dipakai
5	100	0	Tidak dipakai
6	60	60	Dipakai
7	70	20	Dikapai
8	30	80	Dipakai
9	20	100	Dipakai
10	20	80	Dipakai
11	60	80	Dipakai
12	50	100	Dipakai
13	40	40	Dipakai
14	70	40	Dipakai
15	80	20	Dipakai
16	80	40	Dipakai
17	40	60	Dipakai
18	30	80	Dipakai
19	60	60	Dipakai
20	60	80	Dipakai
21	70	20	Dipakai
22	70	20	Dipakai
23	50	100	Dipakai
24	40	60	Dipakai
25	20	80	Dipakai
26	40	60	Dipakai
27	70	40	Dipakai
28	80	20	Dipakai
29	70	40	Dipakai
30	80	40	Dipakai

Sebagaimana hasil analisis tingkat kesulitan, analisis gabungan antara tingkat kesulitan dan daya beda pada Tabel 5 di atas menunjukkan bahwa, butir soal no 1 sd 5 memang tidak bisa digunakan. Butir 6 sd 30 memiliki kriteria yang ideal karena

hasil analisisnya menunjukkan kriteria tingkat kesulitan dan daya pembeda sekaligus. Dengan demikian, bisa dipastikan bahwa butir soal nomor 5 sd 30 dapat digunakan.

Setelah direkonstruksi lagi untuk digunakan dalam ujicoba yang kedua, butir soal nomor 1 sd 5 dibuang. Sedangkan nomor urut butir soal disesuaikan menjadi nomor 1 sd 25. Butir soal inilah yang dianggap versi akhir hasil pengembangan dan butir soalnya disajikan dalam Lampiran 1.

### ***Hasil Analisis Distraktor***

Analisis distraktor dimaksudkan untuk mengetahui berfungsi tidaknya option yang terdiri dari kunci dan pengecoh. Kriteria analisis distraktor dalam laporan ini mengikuti Brown (1995) yang menjelaskan bahwa option dianggap efektif apabila minimal dipilih oleh satu peserta tes. Hasil analisis selengkapnya disajikan dalam Tabel 6 berikut.

Hasil analisis distraktor pada Tabel 6 di atas menunjukkan bahwa terdapat 4 (16%) item yang memiliki option yang tidak efektif. Item tersebut ialah nomor 7(a), 12 (e), 22 (a) dan 26 (e). Distraktor tersebut dianggap tidak efektif karena tidak ada yang memilih. Dalam laporan ini, distraktor tersebut direvisi untuk digunakan pada ujicoba berikutnya, tetapi tidak dianalisis lagi efektivitasnya.

Tabel 6. Analisis Efektivitas Distraktor

Nomor Item	Option					Keterangan
	A	B	C	D	E	
1	5	3	0	2	0	Tdk dipakai
2	0	2	5	0	3	Tdk dipakai
3	0	0	2	8	0	Tdk dipakai
4	0	0	0	6	4	Tdk dipakai
5	6	2	3	0	1	Tdk dipakai
6	5	2	1	1	1	Dipakai
7	0*	2	1	1	5	A direvisi*
8	1	4	1	2	2	Dipakai
9	1	1	6	1	1	Dipakai
10	5	2	1	1	1	Dipakai
11	1	2	1	5	1	Dipakai
12	1	2	6	1	0*	E direvisi*
13	4	2	2	1	1	Dipakai
14	1	1	1	1	6	Dipakai
15	2	4	1	2	1	Dipakai
16	1	5	2	1	1	Dipakai
17	1	2	4	2	1	Dipakai
18	1	2	1	1	5	Dipakai
19	1	1	1	6	1	Dipakai
20	2	5	1	1	1	Dipakai
21	5	1	2	1	1	Dipakai
22	0*	6	2	1	1	A direvisi*
23	5	2	1	1	1	Dipakai
24	1	1	1	4	3	Dipakai
25	1	2	1	5	1	Dipakai
26	1	2	6	1	0*	E direvisi*
27	1	5	2	1	1	Dipakai
28	1	1	5	2	1	Dipakai
29	6	1	1	1	1	Dipakai
30	1	2	1	1	5	Dipakai

### *Hasil Analisis Validitas*

Analisis validitas dilakukan menggunakan teknik validitas criteria. Skor pertama diambil dari hasil tes pada kelas III-B sebanyak 10 orang yang dipilih secara random dan skor kedua diperoleh dari tes membaca pemahaman yang diberikan pada siswa yang sama yang dilakukan oleh guru kelas. Hasil analisis menunjukkan bahwa nilai hitung mencapai 0.917. Menggunakan kriteria persentase seperti disarankan Djiwandono (1995) hasil ini menunjukkan bahwa butir soal ini memiliki kecocokan sampai 91,7%. Jadi, butir soal ini memiliki tingkat validitas yang sangat tinggi. Hasil selengkapnya disajikan dalam Tabel 7 berikut.

Tabel 7. Analisis Validitas Teknik Kriteria

No.	X	Y	x	y	x <sup>2</sup>	y <sup>2</sup>	xy
1	68	80	- 5.6	- 7.5	31.36	56.25	42
2	72	85	- 9.6	- 12.5	92.16	156.25	120
3	72	80	- 9.6	- 7.5	92.16	56.25	72
4	56	70	6.4	2.5	40.96	6.25	16
5	64	75	- 1.6	- 2.5	2.56	6.25	4
6	72	75	- 9.6	- 2.5	92.16	6.25	24
7	60	70	2.4	2.5	5.76	6.25	6
8	68	70	- 5.6	2.5	31.36	6.25	- 14
9	40	60	22.4	12.5	501.76	156.25	280
10	52	65	10.4	7.5	108.16	56.25	78
	624	725			998.4	512.5	656
	62.4	72.5					

Berdasarkan data pada Tabel 7 di atas, analisis menggunakan rumus Product Moment dilakukan sebagai berikut.

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

$$r_{xy} = \frac{656}{\sqrt{(998.4)(512.5)}}$$

$$r_{xy} = \frac{656}{\sqrt{511680}}$$

$$r_{xy} = \frac{656}{715.318111}$$

$$r_{xy} = 0.917$$

### ***Hasil Analisis Reliabilitas***

Reliabilitas dalam laporan ini diperoleh dengan menggunakan teknik test-retest. Tes yang sama diberikan kepada peserta yang sama dalam waktu yang hampir bersamaan. Dalam ujicoba ini, peneliti menggunakan tes pertama yang diberikan pada siswa kelas III-B sebanyak 10 orang yang dipilih secara random. Ke-10 orang tersebut ialah siswa yang tidak dites untuk diambil datanya dalam analisis validitas. Jarak antara tes pertama dengan tes kedua ialah 30 menit. Agar hasilnya lebih meyakinkan, test-retest ini dibedakan waktu pelaksanaannya dengan tes untuk data validitas.

### ***Analisis Reliabilitas Teknik Test-Retest***

Reliabilitas artinya keajegan atau konsistensi skor yang diperoleh siswa dalam tes yang diujicobakan. Jadi analisis reliabilitas maksudnya untuk mengetahui apakah skor yang diperoleh siswa pada tes pertama memiliki konsistensi dibanding dengan tes kedua. Hasil analisis menggunakan rumus Product Moment menunjukkan r-hitung 0.829. Jika dikalikan dengan 100%, diperoleh tingkat konsistensi sebesar 82,9%. Ini artinya, skor tes pertama memiliki tingkat konsistensi sebesar 82,9% atau sangat tinggi. Hasil lengkap analisis disajikan dalam Tabel 8 berikut.

Tabel 8. Hasil Analisis Reliabilitas Teknik Test Retest.

No.	X	Y	x	y	x <sup>2</sup>	y <sup>2</sup>	xy
1	64	60	0.8	5.2	0.64	27.04	4.16
2	64	60	0.8	5.2	0.64	27.04	4.16
3	72	64	- 7.2	1.2	51.84	1.44	- 8.64
4	68	72	- 3.2	- 6.8	10.24	46.24	21.76
5	68	72	- 3.2	- 6.8	10.24	46.24	21.76
6	72	72	- 7.2	- 6.8	51.84	46.24	48.96
7	72	72	- 7.2	- 6.8	51.84	46.24	48.96
8	60	64	4.8	1.2	23.04	1.44	5.76
9	56	60	8.8	5.2	77.44	27.04	45.76
10	52	56	12.8	9.2	163.84	84.64	117.76
	648	652			441.6	353.6	327.68
	64.8	65.2					

Berdasarkan data pada table di atas, analisis menggunakan rumus product Moment dilakukan sebagai berikut.

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

$$r_{xy} = \frac{327.68}{\sqrt{(441.6)(353.6)}}$$

$$r_{xy} = \frac{327.68}{\sqrt{156149.76}}$$

$$r_{xy} = \frac{327.68}{395.1578924}$$

$$r_{xy} = \mathbf{0.829}$$

### Hasil Analisis SPSS Release 11

Analisis SPSS dimaksudkan untuk memberi konfirmasi sejauh mana analisis manual memiliki kecocokan dengan analisis komputer. Tujuan analisis SPSS ialah untuk mengetahui validitas butir soal dan reliabilitas total. Berbeda dengan analisis manual, analisis menggunakan SPSS hanya berfungsi untuk menetapkan validitas dan reliabilitas. Tingkat kesulitan, daya pembeda, dan efektivitas diskartor tidak tersedia dalam program SPSS ini sehingga tidak dilakukan analisis.

Data pada Tabel 8 menunjukkan jumlah butir soal sebelum ujicoba sebanyak 30 dan peserta ujicoba sebanyak 10 orang. Dari tabel 8 ini diidentifikasi jawaban masing-masing peserta dan hasil analisisnya disajikan pada Tabel 9 dan Tabel 10. Hasil analisis pada Tabel 9 dan Tabel 10 menunjukkan bahwa dari 30 butir, butir soal nomor 1 sd 5 tidak valid. Berturut-turut hasil hitung kelima butir tersebut ialah 0.1080, 0.4974, 0.0000, 0.1080, dan 0.4974. Hasil hitung ini berada di bawah kriteria  $N=10-1$  pada  $p=0.05$  dengan nilai 0.6021 sehingga dinyatakan tidak valid.

Butir soal nomor 6 sd 30 memiliki hasil hitung lebih tinggi dibanding kriteria. Dengan demikian butir soal 6 sd 30 dinyatakan valid. Hasil hitung menggunakan SPSS ternyata mengkonfirmasi hasil hitung manual. Ini artinya baik manual maupun SPSS memberikan hasil yang kurang lebih sama.

Tabel 8. Hasil Analisis Validitas Butir Tes Membaca, SPSS Release 11.

NR	Butir 1	Butir 2	Butir 3	Butir 4	Butir 5	Butir 6	Butir 7	Butir 8
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1
6	0	1	1	0	1	0	0	0
7	0	1	1	0	1	1	1	1
8	0	1	1	0	1	1	1	1
9	1	0	1	1	0	1	1	1
10	1	0	1	1	0	0	0	0

Butir 9	Butir 10	Butir 11	Butir 12	Butir 13	Butir 14	Butir 15	Butir 16	Butir 17
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1	1	0	0	1	1	1	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0

Butir 18	Butir 19	Butir 20	Butir 21	Butir 22	Butir 23	Butir 24	Butir 25	Butir 26
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	1	0	1	0	0	0	1	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0

Butir 25	Butir 26	Butir 27	Butir 28	Butir 29	Butir 30
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
1	0	0	1	0	1
1	1	1	1	1	1
1	1	1	1	1	1
1	1	1	1	1	1
0	0	0	0	0	0

**Tabel 9. Analisis Validitas Butir**

\*\*\*\*\* Method 1 (space saver) will be used for this analysis \*\*\*\*\*

R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   ( A L P H  
A)

		Mean	Std Dev	Cases
1.	BUTIR1	.7000	.4830	10.0
2.	BUTIR2	.8000	.4216	10.0
3.	BUTIR3	1.0000	.0000	10.0
4.	BUTIR4	.7000	.4830	10.0
5.	BUTIR5	.8000	.4216	10.0
6.	BUTIR6	.8000	.4216	10.0
7.	BUTIR7	.8000	.4216	10.0
8.	BUTIR8	.8000	.4216	10.0
9.	BUTIR9	.8000	.4216	10.0
10.	BUTIR10	.9000	.3162	10.0
11.	BUTIR11	.9000	.3162	10.0
12.	BUTIR12	.8000	.4216	10.0
13.	BUTIR13	.8000	.4216	10.0
14.	BUTIR14	.9000	.3162	10.0
15.	BUTIR15	.9000	.3162	10.0
16.	BUTIR16	.7000	.4830	10.0
17.	BUTIR17	.8000	.4216	10.0
18.	BUTIR18	.8000	.4216	10.0
19.	BUTIR19	.9000	.3162	10.0
20.	BUTIR20	.8000	.4216	10.0
21.	BUTIR21	.9000	.3162	10.0
22.	BUTIR22	.8000	.4216	10.0
23.	BUTIR23	.8000	.4216	10.0
24.	BUTIR24	.8000	.4216	10.0
25.	BUTIR25	.9000	.3162	10.0
26.	BUTIR26	.8000	.4216	10.0
27.	BUTIR27	.8000	.4216	10.0
28.	BUTIR28	.9000	.3162	10.0
29.	BUTIR29	.8000	.4216	10.0
30.	BUTIR30	.9000	.3162	10.0

Statistics for	Mean	Variance	Std Dev	N of Variables
SCALE	24.8000	85.0667	9.2232	30

—

Tabel 10. RELIABILITY ANALYSIS - SCALE  
(ALPHA)

Item-total Statistics	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
BUTIR1	24.1000	83.8778	.1080	.9815
BUTIR2	24.0000	81.1111	.4974	.9790
BUTIR3	23.8000	85.0667	.0000	.9794
BUTIR4	24.1000	83.8778	.1080	.9815
BUTIR5	24.0000	81.1111	.4974	.9790
BUTIR6	24.0000	77.7778	.9562	.9765
BUTIR7	24.0000	77.7778	.9562	.9765
BUTIR8	24.0000	77.7778	.9562	.9765
BUTIR9	24.0000	77.7778	.9562	.9765
BUTIR10	23.9000	80.3222	.8194	.9774
BUTIR11	23.9000	80.3222	.8194	.9774
BUTIR12	24.0000	77.7778	.9562	.9765
BUTIR13	24.0000	77.7778	.9562	.9765
BUTIR14	23.9000	80.3222	.8194	.9774
BUTIR15	23.9000	80.3222	.8194	.9774
BUTIR16	24.1000	81.8778	.8194	.9802
BUTIR17	24.0000	77.7778	.9562	.9765
BUTIR18	24.0000	77.7778	.9562	.9765
BUTIR19	23.9000	80.3222	.8194	.9774
BUTIR20	24.0000	77.7778	.9562	.9765
BUTIR21	23.9000	80.3222	.8194	.9774
BUTIR22	24.0000	77.7778	.9562	.9765
BUTIR23	24.0000	77.7778	.9562	.9765
BUTIR24	24.0000	77.7778	.9562	.9765
BUTIR25	23.9000	80.3222	.8194	.9774
BUTIR26	24.0000	77.7778	.9562	.9765
BUTIR27	24.0000	77.7778	.9562	.9765
BUTIR28	23.9000	80.3222	.8194	.9774
BUTIR29	24.0000	77.7778	.9562	.9765
BUTIR30	23.9000	80.3222	.8194	.9774

Reliability Coefficients

N of Cases = 10.0

Alpha = .9783

N of Items = 30

## PEMBAHASAN

Hasil analisis ujicoba tes membaca pemahaman literal ini menunjukkan bahwa dari 30 butir soal yang dianalisis, terdapat 25 butir yang dianggap valid dan bisa digunakan dalam tes yang sebenarnya. Proses pembuktian melalui ujicoba ini menggunakan perpaduan teknik manual dan teknik program SPSS.

Untuk memperoleh hasil ujicoba yang baik, diperlukan beberapa kali ujicoba. Laporan ini menggunakan empat kali tes yang diberikan terhadap siswa yang berbeda. Ujicoba yang pertama diberikan pada 10 orang yang diseleksi secara random. Responden ialah siswa kelas III-A MAN 1 Blitar. Tujuan pemberian ujicoba ini ialah untuk mengetahui kualitas butir soal yang meliputi tiga katagori: tingkat kesulitan butir, daya pembeda, dan efektivitas distraktor.

Pelaksanaan ujicoba pertama ini sesuai dengan saran ndari Brown (1995) dan Djiowandono (1995). Hakikatnya, ujicoba pertama bertujuan untuk mengetahui dan memilih batir-butir soal yang dianggap memenuhi kriteria yang diharapkan yaitu mudah, sedang, dan sulit. Proporsi batir soal dengan katagori tersebut perlu diatur secara cermat menggunakan data empiris hasil ujicoba.

Data hasil analisis koalitas batir soal dalam laporan ini menunjukkan bahwa interpretasi kualitas butir harus melibatkan tingkat kesulitan dan daya pembeda. Kedua kriteria tersebut harus mengacu pada kriteria yang bisa diterima. Untuk butir soal yang menggunakan pilihan ganda, analisis distraktor merupakan faktor penting lain yang juga harus diperhatikan. Kajian ini menunjukkan bahwa terdapat empat butir soal yang memiliki distraktor yang harus direvisi. Ini artinya, kecermatan menyusun distraktor juga merupakan hal penting dalam pengembangan butir soal. Analisis tingkat kesulitan dan daya beda menunjukkan analisis internal butir soal.

Dalam analisis validitas, diperlukan teknik yang tepat karena akan menentukan berapa kali ujicoba akan dilakukan. Kajian ini menggunakan analisis validitas kriteria karena dimaksudkan untuk mengetahu validitas eksternal. Teknik ini dianggap memiliki taraf kepercayaan yang lebih lengkap dibanding analisis

internal saja karena kesesuaian dengan kemampuan peserta bisa diidentifikasi dengan cermat.

Masih terkait dengan analisis eksternal, analisis reliabilitas teknik test-retest dianggap memiliki beberapa keuntungan. Pertama, test-retest lebih praktis dilakukan karena data yang digunakan ialah data keras. Melalui dua kali tes yang diberikan pada waktu yang hampir sama terhadap peserta yang sama memberikan kemudahan bagi pengembang tes. Selain itu, analisis menggunakan rumus Product Moment lebih mudah. Namun, teknik ini bisa saja memberatkan peserta karena harus mengerjakan tes dua kali dalam waktu yang hampir sama. Kelemahan ini bisa diatasi dengan cara dipersiapkan sebelumnya dan peserta diberi waktu yang longgar.

Analisis menggunakan program SPSS ternyata lebih praktis dan menunjukkan hasil yang kurang lebih sama dengan hasil analisis manual. Ini artinya, analisis instrumen tidak perlu menggunakan teknik tertentu yang dianggap unggul dan menganggap analisis lainnya kurang baik. Faktor kepraktisan dan kemampuan peneliti merupakan pertimbangan penting untuk mengatasi persoalan ini.

Analisis SPSS dalam kajian ini menunjukkan bahwa hasil akhir yang kurang lebih sama merupakan indikasi bahwa baik manual maupun SPSS sama baiknya. Hanya saja, analisis menggunakan SPSS tidak bisa melihat lebih teliti tingkat kesulitan butir, daya pembeda, dan efektivitas distraktor. Selain itu, menggunakan SPSS kadang tidak dikuasai oleh peneliti sehingga teknik ini dianggap sulit.

Bisa disimpulkan bahwa pengembangan butir soal membaca pemahaman literal dalam kajian ini menunjukkan hasil yang akurat. Dua analisis sekaligus secara manual dan menggunakan SPSS menunjukkan bahwa kedua analisis ini tidak memberi perbedaan yang mengkhawatirkan secara metode.

## **SIMPULAN**

Analisis pengembangan butir soal membaca pemahaman literal dalam kajian ini menunjukkan 25 butir valid dan 5 butir tidak valid. Analisis dilakukan menggunakan teknik manual dan teknik SPSS. Dari 30 butir yang dikembangkan

berdasarkan 7 teks, terdapat 5 butir dalam teks pertama yang menunjukkan hasil tidak valid sehingga tidak bisa digunakan. Sebaliknya, 25 butir soal yang dikembangkan dari 6 teks memiliki tingkat kesulitan, daya beda, distraktor, validitas, dan reliabilitas yang bisa dipercaya.

Analisis butir di sini sengaja menggunakan dua teknik: manual dan SPSS untuk mengetahui perbedaan hasil analisis menggunakan kedua teknik tersebut. Hasil analisis menunjukkan bahwa baik manual maupun SPSS memberikan hasil yang kurang lebih sama. Yang berbeda ialah dalam teknik manual, analisis lebih bervariasi meliputi tingkat kesulitan butir, daya beda, distraktor, validitas, dan reliabilitas. Dalam teknik SPSS, analisis mengacu pada validitas butir dan reliabilitas total saja. Masing-masing teknik memiliki keunggulan dan kelemahan. Dalam teknik manual, pelaksanaan ujicoba cukup kompleks dan dilaksanakan beberapa kali. Dalam teknik SPSS program yang digunakan terbatas dan tidak semua orang menguasai.

Dalam kaitannya dengan kualitas butir yang dikembangkan di sini bisa disimpulkan bahwa kedua teknik tersebut bisa digunakan semua. Analisis butir secara meyakinkan menunjukkan bahwa ke-25 butir yang dianalisis menunjukkan kualitas yang bagus. Kelemahan muncul sedikit dalam bentuk distraktor pada empat butir soal yang tidak dipilih peserta sehingga perlu direvisi.

#### **DAFTAR RUJUKAN**

- Anastasi A and Urbina S. 1997. *Psychological Testing*. Apple Sadle River, NY: Prentice-Hall International Inc.
- Budiharso, T. 2004. *Prinsip dan Strategi Pengajaran Bahasa*. Surabaya: Penerbit Lutfansyah Mediatama.
- Budiharso, T. 2007. Content Analysis of English Test for National Leaving Exam. *Didaktika*, 8(3):1-12.
- Brown, JD. 1995. *Developing Language Test*. New Jersey: Prentice Hall Inc.
- Burns, PC, Roe, BD, dan Ross, EP. 1996. *Teaching Reading in Today's Elementary School*. Boston: Houghton Mifflin Company.
- Cahyono, BY. 1992/1993. *Teori Skema, Struktur Teks, dan Metakognitif pada Pengajaran Membaca Bahasa Inggris*. Malang: Proyek Operasi dan Perawatan Fasilitas.
- Djaali. 2008. *Pengujian dan Pengukuran dalam Pendidikan*. Jakarta: LIPPI.

- Djiwandono, MS. 1995. *Tes dalam Pengajaran Bahasa*. Bandung: Penerbit ITB Bandung.
- Nurhadi. 1987. *Kapita Selekta Kajian Bahasa, Sastra, dan Pengajarannya*. Malang: FPBS IKIP Malang.
- Sharpe, P. 2000. *How to Prepare for the TOEFL Test*. Jakarta: Binarupa Aksara.
- Syafi'ie, I. 1999. *Pengajaran Membaca di Kelas-Kelas Awal Sekolah Dasar*. Pidato Pengukuhan Guru Besar. Malang: Universitas Negeri Malang.
- Valencia, SW, Hiebert EH, Afflerbavh PP. 1996. *Authentic Reading Assessment: Practices and Possibilities*. Newark, Delaware: International Reading Association.

## TEST ON READING COMPREHENSION

**Hari/Tanggal Tes** : 24 November 2008  
**Waktu Tes** : 90 menit  
**Kelas** : SMA Kelas III

### PETUNJUK

Tes berikut ini dimaksudkan untuk menguji kemampuan membaca. Jumlah soal sebanyak 25 butir dan soal dibuat berdasarkan wacana. Terdapat 6 wacana dalam soal ini dan setiap wacana disertai dengan beberapa butir soal. Bacalah teks berikut ini dan jawablah semua pertanyaan yang ada di bawah setiap teks. Tulislah jawaban Anda di lembar jawaban yang sudah disediakan. Jawaban bisa diberikan dengan cara memberi tanda silang atau melingkari pilihan yang disediakan.

#### Text 1: CASSAVA

Cassava is the staple food of million of people in Africa, Asia and Central America. Its swollen tuberous root can be boiled and mashed or grated to produce a meal, known as 'farinha' in Brazil and 'garri' in Nigeria, which can be cooked in small cakes. The root is also the source of the manufactured commodity, tapioca, a small field planted with cassava can be insurance against famine, because the crop can be left in the ground for two or three years without deterioration of the tubers. But recent finding suggest that cassava may be responsible for birth defects.

It has other serious disadvantages. The tubers consist almost entirely of starch and are particularly low in protein, so dependence on cassava leads to serious malnutrition. To make matters worse, some varieties, when grown under certain conditions of soil and climate, develop a high acid content and become extremely poisonous to people and livestock if eaten raw. These tubers have to be well prepared for consumption by prolonged and repeated boiling.

New danger has emerged over the past few years in Nigeria. Doctors have begun to suspect that cassava, if eaten in large amount during pregnancy, may cause deformities in the developing fetus. There appears to be a correlation between the eating of cassava by pregnant women and the occurrence of various kinds of brain or other neuronal malformations in their babies. (*Taken from Headlight An Extensive Exposure to English Learning for SMA Students. Erlangga*)

1. The topic of the text is ...
  - a. The advantages and disadvantages of cassava
  - b. Cassava, the source of various kinds of meals
  - c. Characteristics of people consuming cassava
  - d. The staple food of people in developing countries
  - e. Malnutrition, the effect of eating cassava
  
2. The paragraph one mostly tells us about ...
  - a. The disadvantage of cassava
  - b. The disease caused by consuming large amount of cassava
  - c. The other name of cassava
  - d. The people in Africa, Asia, and Central America
  - e. The advantage of cassava
  
3. How long the cassava crop could be left in the ground without causing deterioration on its tuber?
  - a. More than three years
  - b. About 24 or 36 months
  - c. Less than 12 months
  - d. Less than two years
  - e. More than 36 months
  
4. People can rely on cassava in times of famine because ...
  - a. Cassava can grow in any climate through out the year
  - b. The tubers can last for more than one year as long as they are not be pulled out
  - c. The tubers won't be deteriorating even if it left in the ground more than one year.
  - d. It is the most important food in may parts of the world
  - e. It can be stored for a long time without being rotten
  
5. Which of the following statements is NOT TRUE according to the text?
  - a. Cassavas should not be eaten because they are poisonous
  - b. Farinha and garri are other names for cassava
  - c. Being low protein, cassava may cause malnutrition
  - d. People only consuming cassava for a long periods of time may endanger their health
  - e. Particular conditions of soil and climate can make cassava poisonous

6. It is likely that doctors will advise pregnant women not to eat too much cassava for fear that ...
  - a. They will be deformed
  - b. Their fetus will not develop
  - c. Their bodies will become crippled
  - d. Their babies may have malformed brains
  - e. Their babies' blood will be poisoned

### **Text 2: MEASLES**

Measles, a childhood disease, has caused suffering to mankind for thousands of years. However, the search for an effective measles vaccine was lasted two hundred years and has finally ended in success. Now, for the first time, measles is a preventable disease. You may ask, "How is this important to children?"

Every year measles kills twice as many Americans as polio now does. More children die from measles than from any other common childhood disease. Also complications of some degree occur including pneumonia and ear disorders. Another after-effect of measles-brain damage is less common, but it can have such serious consequences that it deserves special attention.

Brain damage due to measles sounds like something far away from our experience. In reality, like any other injury, damage to the brain can be very slight or very severe. It is quite possible that we have never seen or heard a child who has severe brain damage - the child would either have died or would be in an institution. However, in medical researches a relationship has been found between measles and such things behavior problems, personality changes and dulling of mental ability. For example, a child may be bad-tempered or a little slow to learn after he has recovered from measles. (*Taken from Headlight An Extensive Exposure to English Learning for SMA Students. Erlangga*)

7. Which of the following is NOT an after-effect of measles?
  - a. Ear disorders
  - b. Personality changes
  - c. Polio
  - d. Slow learning
  - e. Pneumonia

8. How many Americans killed by measles every year?
  - a. Twice as many as polio does
  - b. Less than polio does
  - c. More than polio does
  - d. Two hundred peoples
  - e. No one

### Text 3: FOREST

Forests are important to supply human needs. They are also important for keeping soil fertile and for preventing floods. The leaves, which have fallen on the ground, become a substance, which is called humus. Humus is a fertilizer to the soil. It holds rainwater during the wet season, stores it and then the humus being washed away by the rain. The additional water that comes from the forests in the dry season means that the same land can be cultivated two or three times a year. Thus, the fields can produce more crops.

For many years people have not been obeying the government regulation and have been cutting down the trees carelessly. As a result, we lost thousand hectares of our forests. Such careless practices should be stopped, otherwise the remaining forests will be in great dangers. People who cut the forest trees carelessly think very little of the future danger that we might face without forests, nothing prevents the water from running down to the lower land. If it happens, the rushing water will wash away the soil and the humus into the rivers. As a result the rivers will be full of mud. In the wet season this will cause floods, which will destroy the farmlands and villages. *(Taken from Headlight An Extensive Exposure to English Learning for SMA Students. Erlangga).*

9. What is humus? It is ...
  - a. A kind of rain water
  - b. A dried soil
  - c. The roots of a tree
  - d. Additional water that comes from the forest
  - e. A substance made of rotten leaves
10. What is the function of the humus? It ...
  - a. Holds the soil from rushing down
  - b. Keeps the rain water during the wet season
  - c. Cause floods
  - d. Cause the river full of mud
  - e. Produce more crops

11. Which one below is not the importance of the forest?  
Because the forest can ...
- Gives us more food
  - Produce waste
  - Produce good crop
  - Keep soil fertile and prevent floods
  - Protect farmer from hunger
12. How can rice fields be cultivated more than twice a year?  
When ...
- There is rain the whole year
  - The rivers are always full of water
  - They get additional water
  - They get good fertilizer
  - The river is dry
13. We may conclude that the writer ...
- Suggest to cutting down the forests
  - Wants to tell people to plan trees
  - Shows how the trees being cut down
  - Explain the procedure of cutting the trees
  - Warns the reader of the dangers of cutting the trees carelessly

#### **Text 4: THE DIGESTIVE SYSTEM**

When you're hungry and you smell dinner, your mouth waters. This is the first step in changing the food you eat into the kind of food that cells can use. This change is called digestion, and it begins in mouth.

Suppose you are eating sandwich. Your front teeth cut and tear the sandwich. The molars grind it, saliva, the juice in the mouth, gets everything moist.

Saliva also contains a type of chemical called an enzyme. It starts breaking down the starch of the bread into sugar.

The moist ball of the food is carried to the back of the mouth by tongue. Muscles force the food into to esophagus, behind the windpipe. Other muscles force the food to esophagus. They do their work so well that the food would travel along even if the person were doing a headstand. (Taken from *Headlight An Extensive Exposure to English Learning for SMA Students*. Erlangga).

14. When you eat a piece of meat or an apple, the use of your front teeth is for ... them.
- Testing
  - Grinding
  - Chewing
  - Tearing
  - Breaking
15. What substance makes the food in the mouth moist?
- The juice
  - The saliva
  - The milk
  - The starch
  - The water
16. The tongue carries the moist food to ...
- The back of the mouth
  - The stomach
  - The small intestines
  - The esophagus
  - The windpipe

### **Text 5: OUR BODY**

In some ways our body is like a machine. A machine, such as a car engine, needs energy to do its work. The engine burns fuel, usually petrol. The fuel is combined with oxygen from the air. Energy is released.

There are many likenesses between a machine and the human body. But there are some the differences that make us better than a machine. For example, our body can change food into living material and it can grow. Machine cannot do these things. They cannot feel or see and they cannot think and learn as human beings do. The human body can repair worn-out parts. It can even produce new human beings. There is something very special about the human body. It is alive, but machines are not.

The human body has many kinds of organs. For example, it has the skeleton to hold the body erect. It has muscles to move the body. It has blood to carry food to all parts of the body. Our body has the brain to think with. It has the nervous system to control the muscles through our senses. Our senses are windows in the world about us. We learn everything through our senses. Information comes to us through our eyes, ears, nose, tongue skin and some

other senses. These senses tell us what is going on inside our body. (Taken from *Headlight An Extensive Exposure to English Learning for SMA Students*. Erlangga)

17. One similarity between our body and a machine is that both ...
  - a. Produce new human beings
  - b. Use energy to do their work
  - c. Can repair worn-out parts
  - d. Can feel, see, think and learn
  - e. Can change food into living material
  
18. The function of the skeleton is ...
  - a. To hold the body erect
  - b. To carry food
  - c. To move the body
  - d. To control the muscles
  - e. To give us information
  
19. Which paragraph tells us about parts of our body and their functions?
  - a. Paragraph 1
  - b. Paragraph 2
  - c. Paragraph 1 and 2
  - d. Paragraph 3
  - e. Paragraph 2 and 3
  
20. The main idea of paragraph 1 is about ...
  - a. The function of our body
  - b. The production of new human beings
  - c. The differences between the human body and a machine
  - d. The function of a machine
  - e. The part of our body
  
21. The human body can repair worn-out part because it is alive. The underlined word means ...
  - a. Used
  - b. Injured
  - c. Broken
  - d. Decayed
  - e. Destroyed

**Text 6: BADMINTON**

Badminton is an indoor sport that has become popular since the end of World War II. It is a sport for two or four players. The players use light, long-handled rackets to hit a shuttlecock.

In badminton, a fault occurs when the opponent, in attempt to return the shuttlecock safely, allows it to go into or under net. To hit the shuttlecock out of the playing area is also a fault. Men win a game after collecting 15 points. In women's single 11 points wins a game.

Badminton has been growing very rapidly. It has growth both as a backyard recreation and as a competitive sport. The International Badminton Federation was founded in 1934 with nine nations represented. Now it has members in more than 40 nations, and a world tournament for both men and women is held every two years. (*Taken from Headlight An Extensive Exposure to English Learning for SMA Students. Erlangga*).

22. Which of these sentences is correct based on the text?
- Only badminton is an indoor game
  - Badminton can be played either as an indoor or an outdoor game
  - Badminton can only be played outdoors
  - The players are all men
  - The players must be over 15 years old
23. There is only one correct sentence related to the text
- Instead of a shuttlecock the players may use light balls
  - A player can get a point if he can't return the shuttlecock
  - After collecting 15 points, a male player wins a game
  - Badminton is a kind of sport for more than four players
  - Both men and women players have to collect less than 15 points
24. When does a fault occur?
- When the player can't return the shuttlecock to his opponent
  - When the player returns the shuttlecock to his opponent
  - After the player hits the shuttlecock with his racket
  - While the shuttlecock passes through the net
  - When the player falls on the ground

25. Which of the following is not the international match in badminton?
- a. Two men versus two men
  - b. One woman and one man versus one man and one woman
  - c. Women's single
  - d. Men's single
  - e. One woman versus one man

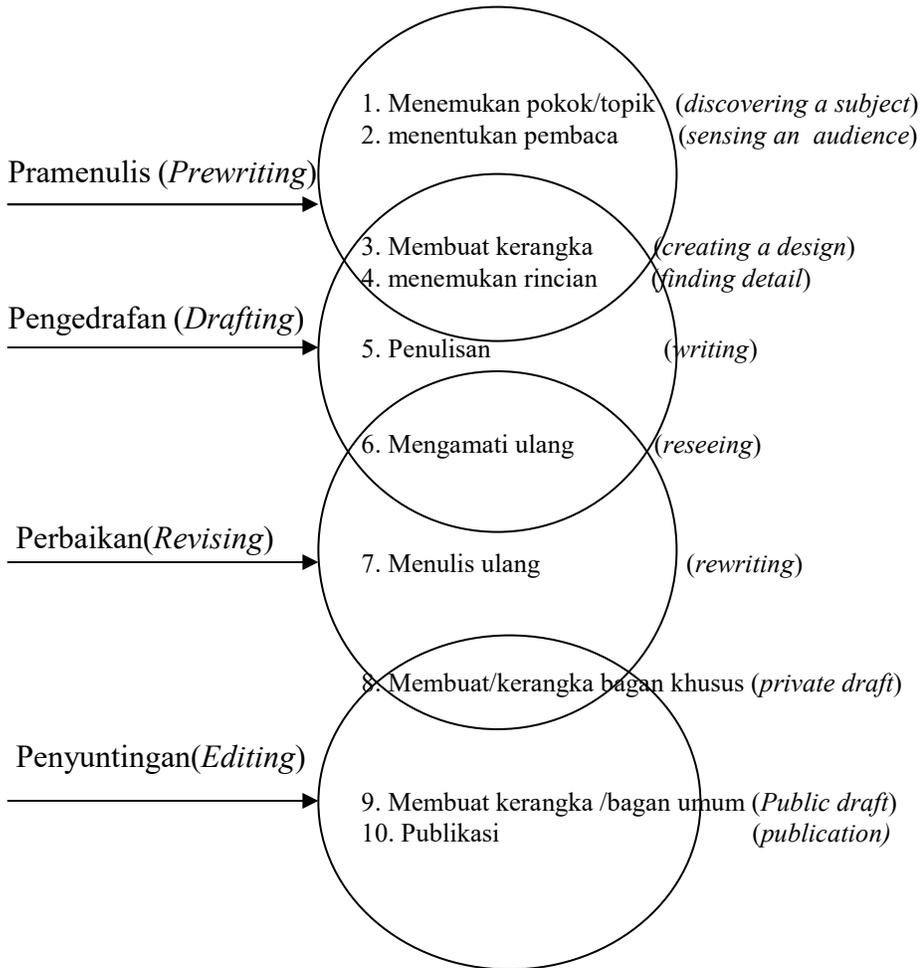
Blitar, 20 November 2008

**BAB 6**  
**CONTOH PENGEMBANGAN TES MENULIS ESAI**

**KERANGKA TEORITIS**

Para peneliti proses menulis membuat tahapan menulis disesuaikan dengan situasi dan kondisi pembelajaran. Ellis dkk.(1989:144) mengungkapkan proses menulis ada empat tahap sebagaimana dikemukakan pada Diagram 1 berikut.

Diagram 1 Proses Menulis Model Ellis, dkk. (1989:144).



Model proses menulis yang tergambar pada diagram di atas menunjukkan bahwa dari tahap satu ke tahap berikutnya dilakukan secara berulang-ulang atau tidak linier. Bila penulis melakukan kegiatan pada tahap berikutnya maka penulis harus melakukan pengecekan kembali terhadap kesesuaian atau keterpaduan antara isi/ makna maupun mekanik yang terkandung pada tahap yang sedang dikerjakan dengan tahap yang telah dilaluinya, demikian seterusnya hingga karangan selesai.

Selain itu, Tompkins (1994:8) membagi proses menulis menjadi tiga tahap kegiatan: tahap konsep, kegiatan untuk memilih dan menentukan topik; tahap inkubasi, yaitu kegiatan pembelajar mengembangkan topik dan menggabungkan informasi yang tersedia; dan tahap hasil, pembelajar melakukan kegiatan mengembangkan tulisan/ menulis kemudian memperbaiki tulisan, dan pada akhir kegiatan menulis pembelajar memeriksa karangan/ tulisan. Tompkins (1994:10) mengemukakan lima tahap berikut.

Tahap 1: Pramenulis (*Prewriting*)

- Pembelajar memilih topik.
- Pembelajar mengumpulkan dan menyesuaikan ide-ide.
- Pembelajar mengidentifikasi pembacanya.
- Pembelajar mengidentifikasi tujuan menulis.
- Pembelajar memilih bentuk yang sesuai berdasarkan pembaca dan tujuan menulis.

Tahap 2: Pengedrafan (*Drafting*)

- Pembelajar menulis draf kasar.
- Pembelajar menulis pokok-pokok yang menarik pembaca.
- Pembelajar lebih menekankan isi daripada mekanik.

Tahap 3: Merevisi (*Revising*)

- Pembelajar membagi tulisannya kepada kelompok.
- Pembelajar mendiskusikan tulisan dengan temannya.
- Pembelajar membuat perbaikan sesuai komentar teman dan gurunya.
- Pembelajar membuat perubahan substantif dan bukan sekedar perubahan minor antara draf pertama dan kedua.

Tahap 4: Mengedit (*Editing*)

Pembelajar membaca ulang tulisannya.  
 Pembelajar membantu baca ulang tulisan temannya.  
 Pembelajar mengidentifikasi kesalahan mekanik dan membetulkannya.

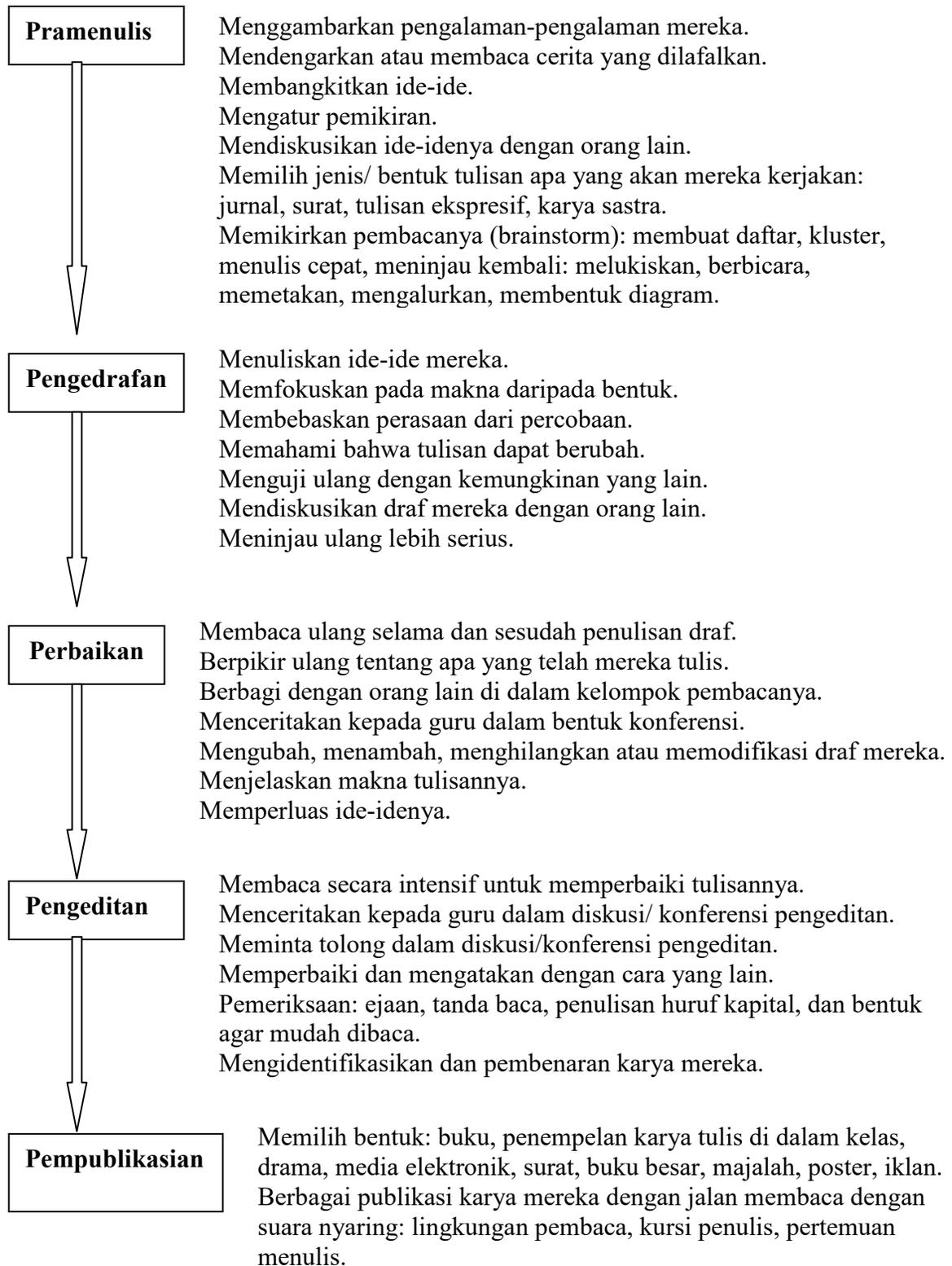
Tahap 5: Mempublikasikan (*Publishing*)

Pembelajar mempublikasikan tulisannya dalam bentuk yang sesuai.  
 Pembelajar membagi tulisannya yang sudah selesai kepada teman sekelasnya.

Burns, dkk. (1996:386) juga mengemukakan bahwa langkah-langkah menulis meliputi lima tahapan: pertama, pramenulis (*prewriting*) dengan aktivitas pengarang persiapan menulis cerita, menggambar, membaca, memikirkan tulisan, menyusun gagasan, dan mengembangkan rencana; kedua, pembuatan draf (*drafting*), dengan aktivitas pengarang merangkaikan gagasan dalam sebuah tulisan tanpa memperhatikan kerapian atau mekanik; ketiga, perevisian (*revising*), pada tahap ini, setelah mendapat saran-saran dari orang lain, pengarang dapat membuat beberapa perubahan, dan perubahan itu dapat melibatkan orang lain; keempat, pengeditan (*editing*), pada tahap ini, pengarang secara hati-hati mengoreksi dan membetulkan ejaan dan mekanisme tulisan; kelima, *sharing* dan publikasi (*sharing and publishing*), pada tahap ini, hasil tulisan bisa dipajangkan di kelas atau dijadikan bahan pustaka di sekolah.

Cox (1999:308) memaparkan tentang menulis sebagai suatu proses dengan lima tahap, yaitu: pramenulis (*prewrite*), pengedrafan (*draft*), revisi/ perbaikan (*revise*), pengeditan (*edit*), publikasi (*publish*). Hal-hal yang dilakukan pembelajar pada tiap-tiap tahap dapat kita lihat pada diagram 2 berikut.

Diagram 2: Proses Menulis Model Cox



Selain membuat tahapan menulis menjadi lima tahap, Tompkins (1994:8) juga membagi proses menulis menjadi tiga tahapan. Pertama, tahap sebelum menulis (*prewriting*). Dalam tahap ini pembelajar melakukan kegiatan memilih topik dan menggabungkan ide. Kedua, tahap menulis (*composing*). Pada tahap ini, pembelajar melakukan kegiatan menulis. Ketiga, tahap setelah menulis (*postwriting*). Pada tahap ini pembelajar melakukan kegiatan memberikan karangan kepada temannya untuk dibaca dan diberikan masukan/ saran.

Keenam pentahapan proses pembelajaran menulis yang dikemukakan oleh Ellis, dkk.(1989), Tompkins (1994), Burns, dkk. (1996), dan Cox (1999), di atas sebenarnya secara umum relatif sama. Pertama, tahap persiapan, kedua, tahap pengembangan tulisan, dan ketiga, perbaikan. Pada tahap persiapan, Britton menyebutkan dengan tahap konsep, sedangkan Burn, dkk. dan Graves menyebut dengan pramenulis. Pada tahap pengembangan tulisan, Britton, menggunakan istilah tahap inkubasi, sedangkan Burn, dkk. dan Grave dengan istilah tahap menulis. Pada tahap perbaikan, mereka agak berbeda-beda. Perbedaan tersebut menunjukkan bahwa untuk membuat karangan dengan baik, banyak cara dan langkah yang dapat digunakan oleh penulis. Selain itu, pentahapan dengan berbagai kegiatan tersebut juga tidak bersifat kaku atau dapat dilakukan berbagai penyesuaian dengan situasi dan kondisi pembelajaran yang dilakukan.

Perbedaan pentahapan dan kegiatan dalam pembelajaran menulis tersebut disebabkan oleh pertimbangan perlu atau tidaknya suatu kegiatan pembelajaran menulis itu dilakukan pada tahap tersendiri atau dapat digabungkan dengan kegiatan menulis lainnya pada tahap yang sama. Hal itu terkait banyak faktor. Misalnya, luas atau tidaknya topik yang ditulis; situasi kelas saat pembelajaran menulis berlangsung pada panas, dingin atau sejuk suhu udara dalam kelas pada siang atau siang hari; dan jumlah pembelajar dalam kelas yang penuh, sedang, atau kurang; kondisi mental (fisik/ psikis) pembelajar dalam menulis, seberapa banyak masalah yang sedang dihadapi pembelajar dalam menulis, seberapa lama pembelajar dapat melakukan

kegiatan menulis; di kelas mana pembelajaran menulis tersebut dilakukan, dengan tujuan apa kegiatan menulis tersebut dilakukan, dan sebagainya.

Pentahapan menulis yang digunakan dalam kajian ini adalah menurut Tompkins dengan lima langkah. Dipilihnya model pentahapan menulis Tompkins (1994:10) untuk proses pembelajaran dalam penelitian ini karena kegiatan-kegiatannya lebih sederhana dan relatif sesuai dengan situasi dan kondisi pembelajar maupun sekolah. Dalam penerapannya, akan diadakan penyesuaian langkah-langkah yang ada dengan situasi dan kondisi anak maupun sekolah yang diteliti.

### **DEFINISI KONSEPTUAL**

Pendekatan proses dalam menulis esai ialah tahapan menuangkan gagasan tertulis dalam bentuk esai dengan menerapkan lima proses secara terintegrasi mulai dari menemukan ide, menyusun draft awal, mengedit dan merevisi naskah sehingga diperoleh naskah akhir berupa esai.

### **DEFINISI OPERASIONAL**

Proses menulis esai ialah perilaku belajar menuangkan gagasan tertulis dalam bentuk esai melalui empat tahapan berupa pra-menulis, pengedrafan, pengeditan, dan revisi yang dapat diukur melalui proses penuangan gagasan berupa menggambarkan pengalaman, mendengarkan atau membaca cerita yang dilafalkan, membangkitkan ide-ide, mengatur pemikiran, mendiskusikan ide dengan orang lain, memilih jenis/ bentuk tulisan apa yang akan mereka kerjakan: jurnal, surat, tulisan ekspresif, karya sastra, memikirkan pembaca (*brainstorm*): membuat daftar, kluster, menulis cepat, meninjau kembali: melukiskan, berbicara, memetakan, mengalurkan, membentuk diagram, menuliskan ide, memfokuskan pada makna, membebaskan perasaan dari percobaan, mendiskusikan draf dengan orang lain, membaca ulang selama dan sesudah penulisan draf, berpikir ulang tentang apa yang telah mereka tulis, menceritakan draft dengan orang lain, mengubah, menambah, menghilangkan atau memodifikasi draf, memperluas ide-idenya, membaca secara intensif untuk

memperbaiki tulisan, memperbaiki dan mengatakan dengan cara yang lain, pemeriksaan: ejaan, tanda baca, penulisan huruf kapital, dan bentuk agar mudah dibaca, mengidentifikasi dan membenaran tulisan, memilih bentuk publikasi.

### KISI-KISI INSTRUMEN

Variabel	Indikator	Kode	Positif	Negatif	+	-	Total
Proses menulis	Pra-menulis	A	1-4	5-7	4	3	7
	Drafting	B	8-10	11-14	3	4	7
	Editing	C	15-18	19-21	4	3	7
	Revising	D	22-24	25-27	3	3	6
	Publikasi	E	28,29	30	2	1	3
					16	14	30

### BUTIR-BUTIR INSTRUMEN ANGKET

- Sebelum mulai menulis, saya menemukan ide dengan cara menggambarkan pengalaman-pengalaman saya:
 

A. selalu	D. jarang
B. Sering	E. tidak pernah
C. kadang-kadang	
- Saya mendengarkan atau membaca cerita atau topik sejenis ebelum saya memulai menulis.
 

A. selalu	D. jarang
B. Sering	E. tidak pernah
C. kadang-kadang	

3. Sebelum mulai menulis, saya berusaha keras untuk membangkitkan ide-ide mengenai topik yang akan saya tulis.  
A. selalu  
B. Sering  
C. kadang-kadang  
D. jarang  
E. tidak pernah
4. Sebelum mulai menulis, saya mengatur pemikiran saya untuk mengurutkan topik dan organisasi ide yang akan saya tulis.  
A. selalu  
B. Sering  
C. kadang-kadang  
D. jarang  
E. tidak pernah
5. Sebelum mulai menulis, saya menghindari untuk mendiskusikan ide-ide saya dengan orang lain.  
A. selalu  
B. Sering  
C. kadang-kadang  
D. jarang  
E. tidak pernah
6. Sebelum mulai menulis, saya merasa tidak perlu memilih jenis tulisan apa yang akan saya gunakan untuk menyampaikan ide.  
A. selalu  
B. Sering  
C. kadang-kadang  
D. jarang  
E. tidak pernah
7. Sebelum mulai menulis, saya merasa tidak perlu memperhatikan siapa yang akan membaca tulisan saya nanti.  
A. selalu  
B. Sering  
C. kadang-kadang  
D. jarang  
E. tidak pernah
8. Untuk memulai menuangkan ide yang akan saya tulis, saya menuangkan ide-ide saya dalam bentuk kerangka awal tulisan.  
A. selalu  
B. Sering  
C. kadang-kadang  
D. jarang  
E. tidak pernah

9. Dalam membuat draft awal tulisan, saya menuangkan makna atau isi yang akan saya tulis dibanding mengoreksi kesalahan-kesalahan gramatika.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
10. Dalam membuat draft awal esai, saya secara bebas menuangkan ide-ide sebagai percobaan.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
11. Saya berusaha menulis draft sekali jadi karena tulisan saya kurang menyukai perubahan-perubahan dalam tulisan saya.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
12. Setelah draft saya selesaikan, saya berusaha tidak memikirkan alternatif lain untuk menuangkan ide yang lebih bagus.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
13. Draft yang sudah selesai saya buat tidak perlu saya diskusikan dengan orang lain untuk menemukan kesalahan atau kekurangan yang mungkin timbul.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
14. Draft yang saya tulis bersifat final dan saya tidak akan meninjau ulang tulisan tersebut secara lebih teliti.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah

15. Untuk memperoleh hasil yang bagus, selama proses menulis draft saya membaca ulang draft tersebut.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
16. Draft yang sudah selesai saya tulis saya renungkan kembali untuk memperoleh penyempurnaan ide.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
17. Saya meminta teman atau orang lain untuk membaca tulisan saya untuk berbagi ide mengenai tulisan tersebut.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
18. Saya mencari masukan dan koreksi mengenai isi tulisan yang sudah saya buat dengan cara diskusi dengan guru atau membahas isi tulisan dengan teman sejawat.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
19. Tulisan saya yang sudah saya tuangkan tidak perlu saya tambah, saya kurangi atau saya modifikasi karena saya berusaha menulis yang bagus sejak awal.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
20. Saya tidak membuat penjelasan mengenai isi dalam tulisan saya karena saya measa sudah cukup lengkap atau tidak perlu saya jelaskan lagi.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
20. Draft tulisan yang saya kembangkan saya usahakan selengkap mungkin sehingga saya merasa tidak perlu memperluas ide-ide dalam tulisan.
- A. selalu
  - B. Sering
  - D. kadang-kadang
  - D. jarang
  - E. tidak pernah

21. Bagi saya, setelah menulis draft yang terpenting ialah kelengkapan isinya, tidak diperlukan lagi revisi atau penuangan ide yang lebih lengkap lagi.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
22. Untuk memperoleh hasil yang baik dari segi kalimat dan penuangan ide, saya membaca secara intensif draft saya untuk bahan perbaikan.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
23. Saya melakukan pengeditan terhadap tulisan yang sudah saya buat dari segi isi, kata, kalimat, dan lain-lain dalam bentuk diskusi atau konferensi pengeditan.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
24. Agar proses pengeditan yang saya lakukan lengkap dan bisa menyempurnakan dari segi kalimat, penggunaan tanda-tanda baca, dan isi tulisan, saya meminta tolong teman atau melakukan koreksi dalam bentuk diskusi.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
25. Jika saya menemukan kata-kata yang salah atau ungkapan yang kurang cocok, saya tidak mengubah dengan cara lain karena tulisan saya sudah final. .
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah
26. Saya tidak mengadakan pengecekan ulang untuk memperbaiki kesalahan-kesalahan ejaan, tanda baca, atau penggunaan huruf kapital karena saya menulis atau mengetik sudah cukup teliti sejak awal.
- A. selalu
  - B. Sering
  - C. kadang-kadang
  - D. jarang
  - E. tidak pernah

27. Sebelum tulisan saya anggap selesai, saya merasa tidak perlu mengidentifikasi kesalahan-kesalahan yang ada dalam tulisan karena tulisan saya sudah benar.
- A. selalu  
B. Sering  
C. kadang-kadang
- D. jarang  
E. tidak pernah
28. Saya mempertimbangkan dengan cermat tulisan yang saya buat akan saya publikasikan dalam bentuk tertentu, misalnya untuk makalah seminar atau jurnal ilmiah, sejak awal menulis sampai akhir tulisan selesai.
- A. selalu  
B. Sering  
C. kadang-kadang
- D. jarang  
E. tidak pernah
29. Saya memperhatikan dengan cermat gaya tulisan dan isi tulisan sesuai dengan jenis publikasi seperti artikel koran, makalah tugas matakuliah, dsb.
- A. selalu  
B. Sering  
C. kadang-kadang
- D. jarang  
E. tidak pernah
30. Untuk membuat tulisan, saya sudah cukup mengerti siapa pembaca tulisan saya, apa tujuan saya menulis dan gaya bahasa yang sesuai dan saya mengusahakan kesempurnaan sejak awal sampai akhir.
- A. selalu  
B. Sering  
C. kadang-kadang
- D. jarang  
E. tidak pernah

## DAFTAR RUJUKAN

- Akhadiah S, dkk. 1997. *Menulis I*. Jakarta: Universitas Terbuka.
- Broto, A.S. 1982. *Metodologi: Proses Belajar Mengajar Berbahasa*. Solo: Tiga Serangkai.
- Burns, P.C., Roe, B.D., & Ross, E. P. 1996. *Teaching Reading in Today's Elementary Schools*. Boston: Houghton Mifflin.
- Cox, C. 1999. *Teaching Language Arts: A Student and Response-Centered Classroom*. Boston: Allyn and Bacon.
- Djaali dan Muljono, Puji. 2008. *Pengukuran dalam Bidang Pendidikan*. Jakarta: Penerbit Grasiono.

- Ellis, A., Panunu, J., Standal, T., & Rummel, M.K. 1989. *Elementary Language Arts Instruction*. Englewood Cliffs, New Jersey: Prentice Hall.
- Tompkins, G. E. 1994. *Teaching Writing: Balancing Process and Product*. New York: Macmillan.
- Tompkins, G.E., & Hoskisson, K. 1991. *Language Arts: Content and Teaching Strategies*. New York: Macmillan.
- Yunus, M. 1997. Hakikat Menulis. dalam Akhadiah, S. (Ed.). *Menulis I*. Jakarta: Universitas Terbuka. Halaman. 1.1-3.45.

### Hasil Analisis Validitas Item Angket Program SPSS 9.30

NR	Butir 1	Butir 2	Butir 3	Butir 4	Butir 5	Butir 6	Butir 7	Butir 8	Butir 9	Butir 10
1	5	5	5	5	5	5	5	5	4	5
2	4	5	4	5	4	5	5	4	5	5
3	5	5	5	5	5	5	4	5	5	5
4	5	4	5	5	5	5	5	5	5	5
5	4	4	4	4	4	5	4	4	4	5
6	4	4	4	4	4	5	4	4	4	5
7	4	4	4	4	4	5	4	4	4	5
8	3	3	5	3	3	5	3	3	3	5
9	5	5	5	5	5	5	5	5	5	5
10	5	5	4	4	5	4	5	5	5	5

### Reliability

\*\*\*\*\* Method 1 (space saver) will be used for this analysis \*\*\*\*\*  
 R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   ( A L P H  
 A)

		Mean	Std Dev	Cases
1.	BUTIR1	4.4000	.6992	10.0
2.	BUTIR2	4.4000	.6992	10.0
3.	BUTIR3	4.5000	.5270	10.0
4.	BUTIR4	4.4000	.6992	10.0
5.	BUTIR5	4.4000	.6992	10.0
6.	BUTIR6	4.9000	.3162	10.0
7.	BUTIR7	4.4000	.6992	10.0
8.	BUTIR8	4.4000	.6992	10.0
9.	BUTIR9	4.4000	.6992	10.0
10.	BUTIR10	5.0000	.0000	10.0

N of

Statistics for	Mean	Variance	Std Dev	Variables
SCALE	45.2000	20.6222	4.5412	10

R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   ( A L P H  
 A)

#### Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
BUTIR1	40.8000	15.0667	.9334	.8778
BUTIR2	40.8000	15.7333	.7932	.8882
BUTIR3	40.7000	19.3444	.2157	.9202
BUTIR4	40.8000	15.2889	.8860	.8814
BUTIR5	40.8000	15.0667	.9334	.8778
BUTIR6	40.3000	21.1222	-.2064	.9279
BUTIR7	40.8000	15.7333	.7932	.8882

BUTIR8	40.8000	15.0667	.9334	.8778
BUTIR9	40.8000	15.7333	.7932	.8882
BUTIR10	40.2000	20.6222	.0000	.9177

Reliability Coefficients

N of Cases = 10.0

N of Items = 10

Alpha = .9064 (reliabilitas total)

## REFERENCES

- Azar, S. Betty. 1992. *Fundamentals of English Grammar*. Englewood Cliffs, NY: Regents Prentice Hall Inc.
- Burke, Jim & Smith, Rick. 2013. *Academic Vocabulary List*. Available online at [www.englishcompanion.com](http://www.englishcompanion.com).
- Finocchiaro & Lavenda. 1970. *Selections for Developing English Language Skills*. New Jersey: Regents Publishing Company Inc.
- Herrera, Mario & Pinkley, Diane. 2012. *Backpack: Workbook Gold 6*. Edinburgh, Essex: Pearson Education.
- Herrera, Mario & Pinkley, Diane. 2012. *Backpack Gold 6*. Edinburgh, Essex: Pearson Education.
- Herrera, Mario & Pinkley, Diane. 2010. *Backpack Gold 5*. Edinburgh, Essex: Pearson Education.
- Hughes, John. 2006. *Telephone English*. Oxford: Macmillan Publishers Limited.
- Joseph, Mukoroli. 2011. *Effective Vocabulary Teaching Strategies for the English for Academic Purposes ESL Classroom*. MA TESOL Collection. Paper 501, pp. 47 to 55.
- Lougheed, Lin. 2005. *How to Prepare for the TOEIC Bridge Test*. Jakarta: PT Binarupa Aksara.
- Mascull, Bill. 2004. *Business Vocabulary in Use: Advanced*. Cambridge: Cambridge University Press.
- McCarthy, Michael & O'Dell, Felicity. 2001. *English Vocabulary in Use: Advanced*. Cambridge: Cambridge University Press.
- McCarthy, Michael & O'Dell, Felicity. 1999. *English Vocabulary in Use: Elementary*. Cambridge: Cambridge University Press.
- Nation, I.S.P. 2009. *Teaching ESL/EFL Reading and Writing*. London: Routledge Publishers Inc.
- Oxford University. 1986. *Reading and Thinking in English: Concept in Use*. Oxford: Oxford University Press.

- Oxford University. 1987. *Reading and Thinking in English: Discovering Discourse*. Oxford: Oxford University Press.
- Oxford University. 1982. *Reading and Thinking in English: Discourse in Action*. Oxford: Oxford University Press.
- Oshima, Alice & Hogue, Ann. 1999. *Writing Academic English*. White Plains, New York: Addison Wesley Longman.
- Oshima, Alice & Hogue, Ann. 2007. *Introduction to Academic English*. White Plains, New York: Addison Wesley Longman.
- Porter, David. 2007. *Check Your Vocabulary for Academic English*. London: A&C Black Publishers.
- Sharpe, J. Pamela. 2000. *Barron's How to Prepare for the TOEFL*. Jakarta: PT Binarupa Aksara.
- Suter, Joanne. 2004. *Building Vocabulary Skills & Strategies Plus Reproducible Activities Level 3*. Rivine, California: Saddleback Educational Publishers.
- West. 1953. *General Service List of English Words*. Available Online.