

SISTEM PENGELOLAAN TES PSIKOLOGI

Oleh : Badrun Kartowagiran (2004)

A. PENDAHULUAN

Menurut tulisan Delandshere dan Petrosky dalam jurnal *The Educational Reseacher*, 27, 2, 1998, teori pengukuran kontemporer yang pada akhir abad 19 sebagai suatu cabang dari ilmu murni, dikembangkan melalui penentuan seperangkat aksioma dan fungsi transformasi angka untuk menterjemahkan dan memformalkan hubungan-hubungan empirik dengan menggunakan angka. Selanjutnya, teori pengukuran berkaitan dengan investigasi sifat dasar dari atribut-atribut fisik dan psikis dasar tertentu.

Secara jelas Campbell yang dikutip Guilford (1954) mengatakan: *measurement as the assignment of numerals to objects or events according to rules*. Sama dengan Campbell, Keeves dan Masters (1999) juga mengatakan bahwa pengukuran adalah pemberian angka (kuantitas numerik) pada obyek-obyek atau kejadian-kejadian menurut aturan. Senada dengan ahli lainnya, Kerlinger (1986) mengatakan bahwa pengukuran adalah pemberian angka pada obyek-obyek atau kejadian-kejadian menurut sesuatu aturan. Sementara itu, Nunnally (1978) menjelaskan bahwa pengukuran itu terdiri dari aturan-aturan untuk memberikan angka/bilangan kepada obyek dengan cara yang sedemikian rupa sehingga dapat mempresentasikan secara kuantitatif sifat-sifat obyek tersebut.

Definisi pengukuran yang dijelaskan para ahli di atas menegaskan bahwa dalam pemberian angka pada subyek, obyek atau kejadian tidak asal memberi angka namun harus menggunakan aturan-aturan, tidak sembarangan. Artinya, orang yang akan memberi angka pada subyek, obyek, ataupun kejadian harus memperhatikan kaidah-kaidah tertentu agar angka yang diberikan sesuai dengan keadaan yang sebenarnya. Semakin jauh seseorang meninggalkan aturan-aturan pengukuran maka semakin besar kesalahan yang terjadi.

Pada makalah ini yang dimaksud dengan pengukuran adalah pengukuran psikologi, bukan pengukuran fisik, atau obyek dan kejadian lainnya. Sehubungan dengan pengukuran psikologi ini, Nunnally (1978) menjelaskan bahwa ada dua tipe kesalahan pengukuran, yaitu: kesalahan sistematik dan kesalahan acak. Kesalahan sistematik terjadi

manakala kualitas instrumen atau alat ukur yang digunakan kurang baik. Sedangkan kesalahan acak dapat disebabkan oleh kondisi subyek yang dites, dan cara menyelenggarakan tes termasuk di dalamnya pelaksana, waktu, dan tempat tes.

Penjelasan tentang sumber kesalahan pengukuran di atas memberi gambaran bahwa kualitas alat ukur merupakan faktor utama dan sangat penting artinya bagi pengukuran. Oleh karena itu perlu dikaji lebih mendalam sistem pengelolaan alat ukur psikologi ini. Menurut Suryabrata (1999), alat ukur psikologi itu dapat dikelompokkan menjadi dua, yaitu: alat ukur yang berkaitan dengan kognitif dan alat ukur yang berkaitan dengan non kognitif. Selanjutnya oleh Azwar (1999) dijelaskan bahwa alat ukur yang berkaitan dengan kemampuan kognitif itu disebut dengan tes. Lebih rinci Suryabrata (1984) menjelaskan bahwa inti dari suatu tes adalah : (1) tes berisi tugas atau serangkaian tugas yang berbentuk pertanyaan-pertanyaan atau perintah-perintah, (2) tes diberikan kepada peserta tes (seseorang atau lebih), dan (3) dilakukan perbandingan tingkah laku antara peserta tes dengan sesuatu yang standar atau dengan tingkah laku peserta lainnya.

Selaras dengan penjelasan di atas dan juga sesuai dengan tugas yang harus diselesaikan oleh penulis, makalah ini hanya membatasi pada sistem pengelolaan alat ukur psikologi aspek kognitif atau sistem pengelolaan tes psikologi. Namun, tidak semua sistem pengelolaan tes psikologi dapat dipaparkan dalam makalah ini. Jelasnya, makalah ini hanya akan memaparkan beberapa sistem pengelolaan tes psikologi, yaitu : sistem pengelolaan soal EBTANAS, tes TOEFL yang dikelola oleh ETS, dan beberapa tes produksi ACT.

B. SISTEM PENGELOLAAN SOAL EBTANAS

1. Pengertian EBTANAS

Pada saat ini penyelenggaraan evaluasi pendidikan secara nasional di Indonesia telah diatur melalui suatu Undang-undang, yaitu UU NO. 2 tahun 1989. Lebih jelasnya termuat dalam Bab XII tentang penilaian. Dalam Bab XII ini secara terperinci termuat langkah-langkah evaluasi dalam pendidikan yang harus dilaksanakan, tanpa memandang jenjang pendidikannya. Untuk jenjang pendidikan dasar dan menengah, evaluasi akhir suatu program dilakukan setiap empat bulan sekali, yang biasa disebut dengan Ulangan Umum. Pelaksanaan evaluasi semacam ini didasarkan pada sistem yang diterapkan dalam pendidikan dasar dan menengah yang menganut sistem catur wulan. Hal ini sesuai

dengan isi dari pasal 46 (1) yang menyebutkan bahwa dalam rangka pembinaan satuan pendidikan pemerintah melakukan penilaian setiap satuan pendidikan secara berkala.

Pada catur wulan ke tiga, evaluasi dilakukan secara menyeluruh. Hal ini disebabkan hasil evaluasi pada catur wulan ke tiga digunakan untuk menentukan tingkat yang boleh diikuti oleh siswa. Dengan kata lain, hasil evaluasi tersebut digunakan untuk merekomendasikan apakah seorang siswa dapat melanjutkan ke kelas atau jenjang pendidikan yang lebih tinggi atau tidak. Khusus untuk tahap akhir dari suatu jenjang pendidikan, evaluasi diselenggarakan secara tersendiri, dan ada pula yang bertaraf nasional. Untuk akhir jenjang pendidikan dasar dan menengah, yakni untuk kelas 6 SD, kelas 3 SLTP, dan kelas 3 SLTA dilakukan Evaluasi Belajar Tahap Akhir (EBTA), dan Evaluasi Belajar Tahap Akhir Nasional (EBTANAS).

EBTANAS dilakukan pertama kali pada akhir tahun ajaran 1980/1981. Sebagai tahap perintisan, mata pelajaran atau bidang studi yang diujikan adalah Pendidikan Moral Pancasila (PMP) dan hanya dilakukan oleh sekolah yang bersedia. Pada tahun ajaran berikutnya (1981/1982), mata pelajaran yang diujikan dalam EBTANAS adalah PMP dan Bahasa Indonesia, sedangkan mata pelajaran lainnya diujikan dalam EBTA. Penyelenggaraan EBTANAS berkembang dari tahun ke tahun, baik dari segi perencanaan, penyiapan bahan ujian, jumlah mata pelajaran yang diujikan maupun mekanisme pelaksanaannya di setiap jenjang. Saat ini jumlah mata pelajaran yang di-EBTANAS-kan pada jenjang pendidikan dasar dan menengah dapat dilihat pada Tabel 1.

Tabel 1. Mata Pelajaran yang Di-EBTANAS-kan

<p>Sekolah Dasar (SD) dan atau yang sederajat:</p> <ol style="list-style-type: none"> 1. PMP 2. Bahasa Indonesia 3. Matematika 4. IPA 5. IPS 	<p>Sekolah Lanjutan Tingkat Pertama (SLTP) dan atau yang sederajat:</p> <ol style="list-style-type: none"> 1. PMP 2. Bahasa Indonesia 3. Matematika 4. IPA 5. IPS 6. Bahasa Inggris
--	--

bersambung

sambungan

Sekolah Menengah Umum (SMU) IPA dan atau yang sederajat : 1. PMP 2. Bahasa Indonesia 3. Matematika 4. Fisika 5. Biologi 6. Kimia 7. Bahasa Inggris	Sekolah Menengah Umum (SMU) IPS dan atau yang sederajat : 1. PMP 2. Bahasa Indonesia 3. Matematika 4. Ekonomi 5. Sosiologi 6. Tata Negara 7. Bahasa Inggris
Sekolah Menengah Kejuruan (SMK) : 1. PMP 2. Bahasa Indonesia 3. Matematika 4. Bahasa Inggris 5. IPA 6. IPS 7. Teori Kejuruan	

Tujuan diselenggarakannya EBTANAS adalah untuk : (1) merintis terciptanya standar nasional mutu pendidikan dasar dan menengah, (2) menyederhanakan prosedur penerimaan siswa baru pada sekolah yang lebih tinggi, (3) mempercepat peningkatan dan pemerataan mutu pendidikan dasar dan menengah, (4) menunjang tercapainya tujuan kurikulum, dan (5) mendorong agar proses belajar-mengajar dilaksanakan berdasar kurikulum, buku, alat peraga yang telah ditetapkan (Depdikbud, 1986). Hal ini ditegaskan kembali oleh Boediono dan Jahja Umar (1999) yang menjelaskan bahwa fungsi utama EBTANAS adalah untuk : (1) sertifikasi, (2) seleksi, dan (3) pemantauan dan pengendalian mutu pendidikan.

2. Sistem Pengelolaan EBTANAS

a. Kepanitiaan

Di setiap tahunnya selalu diterbitkan buku pedoman penyelenggaraan EBTANAS, baik di tingkat pusat, wilayah, rayon, maupun sub-rayon. Pedoman penyelenggaraan EBTANAS tingkat pusat didasarkan pada Surat Keputusan Bersama (SKB) Direktorat Jenderal Pendidikan Dasar dan Menengah (Dirjen Dikdasmen) Departemen Pendidikan dan Kebudayaan (Depdikbud) dan Direktorat Jenderal Pembinaan Kelembagaan Agama Islam (Dirjen Binbagais) Departemen Agama (Depag). Sebagai contohnya SKB nomor 423/C/Kep/PP/1998, Nomor E/344.A/1998 tentang Pedoman Penyelenggaraan EBTANAS tahun pelajaran 1998/1999.

Pedoman penyelenggaraan EBTANAS seperti yang dijelaskan di atas memuat beberapa hal penting, yaitu: (1) pendahuluan, (2) tujuan EBTANAS, (3) penyelenggaraan yang meliputi: penyusunan soal, penggandaan soal, sekolah penyelenggara, jadwal, pengawasan, pemeriksaan, mata pelajaran yang di-EBTANAS-kan, daftar NEM, penyusunan dan penerbitan daftar kolektif NEM khusus untuk SMK, kegunaan NEM, (4) perhitungan nilai untuk penentuan nilai STTB, (5) penentuan keberhasilan khusus untuk SMK, (6) sertifikasi kompetensi khusus untuk SMK, (7) organisasi penyelenggaraan, (8) pemantauan, evaluasi, pelaporan, dan tindak lanjut, (9) pengamanan dan sanksi, dan (10) pembiayaan.

Selanjutnya, berdasarkan SKB ini daerah menyusun pedoman penyelenggaraan EBTANAS yang disesuaikan dengan kondisi daerah masing-masing.

Adanya pedoman yang rinci mendorong pelaksanaan pengelolaan EBTANAS yang mencakup: kepanitiaan, jadwal dan tempat, pengawasan, pengamanan, pembiayaan, dan pemeriksaan dapat berjalan dengan baik. Kepanitiaan terdiri dari panitia tingkat : Pusat/ nasional, Kanwil/wilayah, Kabupaten/Rayon, Kecamatan /sub-rayon (untuk SD/MI), dan sekolah penyelenggara EBTANAS. Kepanitiaan EBTANAS dibentuk setiap tahun, dimulai menjelang dilaksanakan ujian EBTANAS dan diakhiri setelah pengumuman dan pelaporan hasil penyelenggaraan EBTANAS. Diagram kepanitiaan penyelenggaraan EBTANAS dapat diperiksa pada Gambar 1.

Dari gambar 1 dapat diketahui bahwa secara nasional penanggung jawab pelaksanaan EBTANAS ini dilakukan secara bersama-sama antara Dirjen Dikdasmen Depdikbud dan Dirjen Binaga Islam Depag. Dikdasmen Depdikbud (sekarang Diknas) menangani sekolah-sekolah umum, sedangkan Binagais Depag menangani sekolah-sekolah Islam, seperti Madrasah Ibtidaiyah, Madrasah Tsanawiyah, dan Madrasah Aliyah. Untuk tingkat wilayah penanggung jawabnya adalah Kakanwil dan daerah tingkat II penanggung jawabnya Kakandep Dikbud/Diknas.

Gambar 1. Struktur Panitia EBTANAS Bersama Ditjen Dikdasmen Depdikbud dan Ditjen Binaga Islam Depag

b. Bank Soal atau Bank Butir Tes

Menurut hasil penelitian Mardapi dan Kartowagiran (1999), prosedur penulisan soal EBTANAS diawali dengan penulisan kisi-kisi. Guru yang terlatih dan terpilih diundang ke Bagian Pelaksana Teknis Ditjen Dikdasmen (BPTDD) Depdikbud Jakarta untuk diberi penyegaran tentang cara menyusun kisi-kisi yang baik. Selanjutnya, bersama-sama dengan ahli dari Pusiasjian Balitbang Depdikbud para guru tersebut menyusun kisi-kisi. Dalam penyusunan kisi-kisi personil yang terlibat adalah guru-guru atau tenaga kependidikan yang pernah dilatih dan menguasai bidang studi didampingi para ahli dari Puslitbangsisijian. Tujuan penyusunan kisi-kisi adalah untuk memberi arahan tentang tujuan pembelajaran, pokok bahasan, sub-pokok bahasan, indikator, dan soal yang akan disusun. Setelah kisi-kisi tersusun, langkah selanjutnya adalah penulisan butir-butir soal.

Untuk tingkat SD, soal ditulis oleh guru yang sudah terlatih dan terpilih. Penulisan dilakukan dan dikoordinasi oleh Kanwil Depdikbud. Dalam penulisan butir-butir soal ini para guru mengacu pada kisi-kisi yang telah diberikan kepada mereka. Untuk tingkat SLTP, mulai tahun 1998 hampir semua soal EBTANAS diambil dari **bank soal**. Sedangkan soal EBTANAS untuk SLTA (SMU dan SMK) ada yang diambil dari **bank soal** tetapi juga ada yang menggunakan soal hasil karya guru. Soal EBTANAS tingkat SLTA yang tidak diambil dari bank soal Pusiasjian Jakarta adalah soal-soal EBTANAS yang disusun oleh guru terlatih dan terpilih. Penulisan soal ini biasanya diselenggarakan dan dikoordinir oleh Kanwil Depdikbud.

Diharapkan ditahun-tahun mendatang semua soal EBTANAS SLTP dan SLTA, khususnya SMU, akan diambilkan dari bank soal. Dengan demikian bank soal merupakan unsur utama dalam penyediaan soal EBTANAS yang berkualitas. Oleh karena itu, dalam makalah ini akan dijelaskan secara rinci tentang bank soal ini. Penjelasan tentang bank soal ini diambilkan dari buku : *Development and Management of an Item Banking System : A guidebook* yang ditulis oleh Richard Sandman dan Sumadi Suryabrata.

1) Pengertian bank soal atau bank butir

Bank soal atau bank butir adalah suatu kumpulan besar butir-butir test yang mengukur suatu bidang instruksi (pengajaran) tertentu. Butir-butir tersebut dikumpulkan selengkap mungkin agar bisa mewakili kandungan area subyek tersebut. Setiap butir dalam bank tersebut biasanya diklasifikasikan menurut materi yang tercakup oleh butir,

atau tujuan khusus pengajaran yang diujinya. Karena tujuannya adalah hanya memasukan butir dengan kualitas bagus ke dalam bank butir, butir-butir tersebut sering kali diujicobakan sebelum diterima ke dalam bank butir. Dalam hal ini, karakteristik statistik dari butir tersebut juga dimasukan ke dalam bank butir.

Sebagian pembahasan dalam buku panduan ini akan terkait dengan bank butir yang diujikan (**calibrated**). Suatu bank butir yang diujikan terdiri dari butir-butir tes yang memenuhi kriteria statistik tertentu dari Teori Respon Butir. Butir-butir tersebut bisa diberi index kesulitan yang mewakili tingkat kesulitan butir dikaitkan dengan butir lain dalam bank butir. Skala kesulitan butir biasanya berkisar antara -0.5 sampai 0.5 , dengan nol sebagai rata-rata perkiraan untuk butir-butir dalam suatu bank butir. Siswa yang mengikuti suatu ujian yang terdiri dari butir-butir tersebut bisa diberikan suatu index kemampuan, yang didasarkan pada performa tes mereka, yang akan ada pada skala yang sama dengan index kesulitan butir. Index kesulitan butir disebut suatu calibration (pengujian), dan setelah ditentukan, juga dimasukan ke dalam bank butir.

2) Manfaat Bank butir

Ada beberapa manfaat bank soal atau bank butir yang perlu dijelaskan kembali, yaitu:

- a) Bank butir adalah sumber siap-sedia untuk butir-butir ujian. Ini akan menghemat waktu dalam mempersiapkan ujian, karena banyak butir baru mungkin tidak perlu untuk ditulis.
- b) Jika butir-butir dalam bank butir sudah dicek kualitasnya, seperti yang biasanya dilakukan, test-test yang berisi butir-butir ini akan berkualitas tinggi juga.
- c) Jika diperlukan bentuk paralel dari suatu tes, bank butir bisa memberikan sejumlah butir yang memadai untuk membentuk bentuk-bentuk paralel tersebut. Juga, statistik yang ada untuk butir tersebut juga bisa menjamin bahwa test yang dihasilkan akan benar-benar paralel.
- d) Jika bank butir mempunyai sejumlah besar butir, kekhawatiran mengenai keamanan dalam memproduksi dan menggunakan test-test tersebut akan berkurang. Para siswa tidak akan mungkin mengingat butir-butir dalam jumlah yang sangat besar, dan dengan penggunaan bentuk paralel, siswa tidak akan mengetahui bentuk khusus yang kan mereka dapatkan.

e) Jika suatu bank butir teruji (**calibrated**) yang digunakan, test-test yang berbeda bisa dibuat untuk menyamai karakteristik kemampuan dari kelompok yang berbeda. Karena tingkat kesulitan butir dan kemampuan siswa ada pada skala yang sama, skor-skor yang dihasilkan dari test-test yang berbeda tersebut akan bisa setara.

3) Pengertian Sistem Perbankan Butir

Ujian harus dikembangkan untuk banyak area subyek dan tingkat sekolah yang berbeda-beda. Setiap kombinasi subyek dan level tersebut bisa mempunyai bank butirnya masing-masing. Contohnya, bisa ada suatu bank butir untuk Matematika kelas 5 SD. Dan mungkin ada juga bank butir untuk Bahasa Inggris pada kelas 2 SMP. Untuk setiap bank butir yang dibutuhkan, butir-butir harus ditulis, diujicobakan, dan dimasukkan ke dalam bank butir. Test perlu dibuat dari butir-butir ini untuk memenuhi tujuan-tujuan tertentu. Terakhir, bank-bank butir sendiri perlu agar tetap *up-to-date*. Kumpulkan bank butir, bersama dengan keseluruhan proses pengembangannya, penggunaannya untuk membentuk ujian yang dibutuhkan, dan membuatnya *up-to-date*, disebut sistem perbankan butir. Jelas suatu sistem perbankan butir akan digunakan secara maksimal dan efektif dalam pengimplementasian suatu program pengujian.

4) Proses Terbentuknya Suatu Sistem Perbankan Butir

Dalam menyusun suatu sistem bank butir, tentu saja tidak dikerjakan seorang diri. Hal ini dapat dimengerti karena banyak hal yang harus dikerjakan, menghabiskan banyak waktu, dan membutuhkan dana untuk mendukungnya. Oleh karenanya, biasanya dibutuhkan satu kelompok personel pendidikan yang bekerja sama untuk membentuk dan meng-implementasikan suatu sistem perbankan butir, untuk memenuhi kebutuhan pengujian yang dilihat para anggota kelompok sebagai kepentingan bersama. Di Kanwil provinsi, contohnya, dilakukan usaha-usaha untuk menetapkan unit-unit pengujian khusus. Unit-unit pengetesan tersebut akan bertanggung jawab atas semua aktivitas di Kanwil, termasuk operasi dari suatu sistem perbankan butir. Dalam kasus tersebut, personel yang dibutuhkan, tanggung jawab individual mereka, peralatan dan ruang yang dibutuhkan, serta kebutuhan dana dan sumber yang mungkin ada semuanya harus ditentukan dan dinyatakan terlebih dahulu, dan kewenangannya harus sudah diberikan. Tanpa organisasi dan perencanaan seperti itu, tampaknya tidak mungkin tercipta suatu

implementasi dari suatu aktivitas seperti yang dilibatkan dalam operasi sistem perbankan, seperti yang dibahas dalam manual ini.

5) Pengembangan Spesifikasi Tes

Butir-butir test baru untuk suatu bank butir biasanya dikembangkan dengan dikaitkan dengan pemenuhan persyaratan butir untuk suatu test tertentu yang sedang dalam proses pembuatan. Persyaratan-persyaratan butir ini diberikan dalam spesifikasi test untuk test baru tersebut. Jika bank butir pada saat itu tidak mempunyai jumlah butir yang sesuai secara memadai untuk memenuhi kebutuhan yang dinyatakan tersebut, butir-butir tambahan harus disiapkan. Butir-butir baru tersebut tentunya akan didasarkan pada informasi yang terkandung di dalam spesifikasi test tersebut.

Suatu pernyataan spesifikasi test yang jelas akan memungkinkan para penulis butir untuk mempersiapkan butir-butir yang benar-benar memenuhi persyaratan pengukuran, dan hal itu akan memberikan tambahan yang bermanfaat bagi bank butir.

Langkah pertama dalam pengembangan spesifikasi test adalah menentukan informasi identifikasi dasar untuk test baru yang sedang dibuat tersebut. Berikut ini adalah beberapa informasi yang terdapat dalam spesifikasi tes.

a) Informasi Identifikasi Dasar

- Tahun kurikulum
- Level sekolah
- Level kelas
- Bidang subyek
- Program studi
- Tahun kademik
- Kuartal tahun sekolah
- Kelas tertentu yang akan diuji
- Tanggal pengujian yang diperkirakan

Sesudah itu, beberapa keputusan awal mengenai test yang akan dibuat tersebut perlu dibuat. Keputusan-keputusan tersebut antara lain:

b) Keputusan yang perlu dibuat

- Tujuan test
- Bahan yang akan diujikan

- Tipe butir yang akan digunakan
- Level kesulitan butir yang dipilih
- Jumlah total butir untuk test tersebut
- Waktu yang dizinkan untuk mengerjakan test

c) Variabel-variabel distribusi butir

Dalam variabel distribusi butir ini, hal-hal yang perlu diperhatikan adalah:

- Kandungan masalah subyek
- Format butir
- Level fungsi kognitif
- Tujuan Perilaku Khusus

Hal-hal di atas merupakan kegiatan-kegiatan kunci dalam mengembangkan spesifikasi tes. Apabila pengembangan spesifikasi tes ini sudah selesai, maka langkah selanjutnya adalah penulisan butir-butir tes.

6). Penulisan Butir-butir Tes Obyektif

Di dalam masyarakat pengujian, orang tahu bahwa kemampuan untuk menulis butir test lebih bersifat seni dibanding hal yang ilmiah. Namun, pengetahuan dalam masalah teknis yang terkait dengan penulisan butir, dan praktiknya dalam penulisan butir yang menggunakan panduan yang baik, akan meningkatkan kemampuan ini. Berikut ini akan disampaikan beberapa “Panduan Umum” dengan harapan bahwa hal tersebut akan membantu penulisan butir yang baik. Harus diingat bahwa panduan-panduan ini hanyalah rekomendasi bukan aturan. Butir-butir panduan umum itu adalah :

- a) Ungkapkan pertanyaan dengan jelas
- b) Jika mungkin, pilih kata-kata yang mempunyai makna yang pasti
- c) Hindarkan susunan kata yang kaku dan rumit
- d) Masukkan semua informasi yang diperlukan untuk memilih jawaban
- e) Hindarkan penggunaan kata-kata yang tidak fungsional
- f) Rumuskan pertanyaan dengan seakurat mungkin
- g) Sesuaikan level kesulitan dengan kelompok dan tujuan pengujian
- h) Hindari pemberian petunjuk yang tidak diperlukan pada jawaban yang benar.

7) Peninjauan dan Revisi Butir Tes

Setelah butir-butir test ditulis, adalah hal yang penting untuk meninjaunya

kembali. Hal ini dikarenakan sebagian butir baru bisa mengandung kelemahan yang mengurangi kualitas butir tersebut. Jika butir-butir tersebut tidak ditinjau kembali, kelemahan-kelemahan ini mungkin tidak akan dbutirukan. Dalam kasus ini, butir yang jelek akan masuk ke dalam bank butir, dan pada akhirnya akan masuk ke dalam ujian tertentu yang sedang dibuat. Hasilnya akan menjadi pengukuran prestasi siswa yang tidak akurat. Tentu saja, apabila ditemukan butir yang lemah maka butir itu direvisi sehingga semua kelemahannya bisa dihilangkan, dan butir tersebut kualitasnya akan tinggi.

Ada beberapa prinsip dasar dalam peninjauan kembali dan revisi butir, yaitu :

- a) Melaksanakan tinjauan kembali butir sebelum melakukan uji coba pengujian
- b) Merevisi butir segera setelah kelemahannya dbutirukan
- c) Butir-butir harus ditinjau kembali oleh orang yang berbeda dari orang yang sebelumnya menuliskan butir tersebut.
- d) Peninjauan kembali dan revisi butir harus dilaksanakan dengan menggunakan suatu proses kelompok.

Secara rinci dapat dijelaskan bahwa dalam menulis butir test harus diperhatikan hal – hal berikut:

- a) Jawaban dari soal jelas kebenarannya
- b) Semua alternatif lain jelas kesalahannya
- c) Butir soal sesuai dengan masalah dan indikator subyek
- d) Butir sesuai tingkat kesulitannya
- e) Konsep atau proses yang sedang diujikan adalah jelas
- f) Istilah dan situasi yang digunakan dalam butir ditentukan dengan jelas.
- g) Siswa dapat memahami apa yang mereka harus lakukan
- h) Siswa akan mendapatkan jawaban sesuai yang diharapkan
- I) Butir ditulis dengan menggunakan bahasa dan ejaan yang tepat
- j) Informasi yang cukup diberikan untuk menjawab butir tersebut
- k) Informasi yang tidak perlu tidak diberikan
- l) Gambar dan grafik diberikan dengan jelas
- m) Pengecoh jawaban yang diberikan adalah masuk akal
- n) Struktur tata bahasa alternatif jawaban konsisten dan layak
- o) Petunjuk yang mengarahkan siswa ke jawaban yang tepat tidak diberikan

8) Uji Coba Butir-butir Tes

Setelah butir-butir baru ditinjau kembali dan direvisi, langkah selanjutnya adalah uji coba. Uji coba ini dilaksanakan untuk mengecek kualitas butir yang baru, dan untuk mendapatkan informasi statistik mengenai mereka. Prosedur yang umumnya digunakan adalah melaksanakan suatu test yang mengandung butir-butir baru tersebut pada suatu sampel siswa pada level yang sesuai.

a) Manfaat Uji Coba Butir-Butir Baru

- (1) Penguji-cobaan butir-butir test baru akan memungkinkan kita untuk mengecek kualitas butir tersebut.
- (2) Penguji-cobaan butir baru memungkinkan kita mendapatkan statistik yang relevan untuk butir tersebut.

b) Prosedur-Prosedur yang Direkomendasikan Untuk Penerapan Uji Coba

- (1) Buat bentuk test untuk uji coba yang sesuai dalam bahan yang dikandungnya pada level siswa yang ada untuk uji coba dan tahun test tersebut akan diujikan
- (2) Gunakan lebih dari satu bentuk test untuk suatu subyek dan level tertentu, jika diperlukan
- (3) Jika suatu bank butir yang terkalibrasi sedang dikembangkan, gunakan butir penghubung untuk menghubungkan bentuk-bentuk berbeda untuk suatu level dan subyek tertentu.
- (4) Gunakan sampel siswa yang representatif untuk uji coba
- (5) Adalah hal yang lebih baik untuk mengetes satu kelas dalam beberapa sekolah yang berbeda daripada mengetes level kelas yang berbeda dalam sedikit sekolah
- (6) Jika suatu bank butir **calibrated** sedang dikembangkan, dan **calibration** diperlukan, jumlah siswa yang diberi test uji coba minimal 750 orang.
- (7) Dorong siswa untuk meninjau kembali (review) bahan test sebelum uji coba dilaksanakan.
- (8) Jangan adakan uji coba dalam beberapa subyek untuk siswa-siswa kelas yang sama.
- (9) Temukan cara-cara untuk memotivasi siswa untuk mengerjakan soal uji coba dengan

serius dan menggunakan usaha mereka yang terbaik.

(10) Pastikan bahwa siswa mempunyai waktu yang cukup untuk menjawab butir-butir soal dalam uji coba.

9) Penganalisaan Hasil Uji Coba

Kita sekarang telah melaksanakan uji coba butir-butir baru pada siswa kelas yang tepat, dan mempunyai data respon. Sekarang apa yang harus kita lakukan adalah menganalisis data ini dan menentukan butir yang mana yang mempunyai kualitas tinggi, dan sehingga diterima masuk bank butir. Butir-butir yang tersisa, yang masih menunjukkan kelemahan, akan perlu direvisi lebih jauh, atau dihapus sama sekali.

Berikut ini adalah prosedur umum yang digunakan untuk menganalisa hasil uji coba, dan pembahasan di sini hanya akan terkait dengan butir pilihan ganda:

- a) Jalankan ITEMAN, suatu program analisis butir umum yang memberikan informasi statistik dasar pada tiap butir baru
- b) Meninjau kembali semua butir, dengan menggunakan informasi ITEMAN dengan mengacu pada syarat butir yang baik. Butir yang jelek direvisi atau dihilangkan
- c) Jalankan BIGSTEPS (program yang mengecek kesesuaian butir dengan model Rasch Teori Respon Butir, dan memberikan index kesulitan pada tiap butir yang sesuai model) pada butir yang lulus tahap 2 tanpa revisi atau dihapus
- d) Beri tanda lulus masuk bank butir bagi butir yang lulus tahap 2, dan untuk bank butir terkalibrasi, butir yang sesuai dengan model Rasch. Butir-butir ini dan statistiknya akan masuk bank butir yang relevan.

Ada dua prinsip yang harus diperhatikan dalam peninjauan kembali butir tes yang menggunakan hasil ITEMAN, yaitu :

- a) Jangan menolak suatu butir hanya dengan berdasarkan statistik ITEMAN saja
- b) Tinjau kembali semua butir, bahkan butir-butir yang tidak dipilih berdasar hasil ITEMAN

10) Penyimpanan Butir-butir Tes Dalam Suatu Bank Butir

Setelah hasil uji coba dianalisis, dan butir berkualitas tinggi telah ditandai lulus masuk ke dalam bank butir, langkah selanjutnya adalah memasukan butir-butir ini ke dalam bank butir. Ini adalah lokasi penyimpanan untuk butir-butir tersebut, tempat untuk didatangi saat butir-butir test dibutuhkan untuk suatu ujian baru yang sedang dibuat.

Dengan menempatkan semua butir dalam satu tempat yang mudah diambil dibutuhkan.

Ada dua cara yang saat ini digunakan untuk menyimpan butir dalam bank butir. Satu melalui penyimpanan butir pada kartu-kartu, satu butir per-kartu, dan kedua adalah menyimpan butir pada komputer. Keduanya akan dibahas secara bergantian:

a) Penyimpanan Butir Pada Kartu

Ini adalah cara lama penyimpanan bank butir. Setiap butir secara manual dimasukkan ke dalam satu kartu, bersama dengan informasi identifikasi dan statistik terkaitnya. Kartu-kartu ini disimpan dalam suatu file/arsip, yang tentunya, adalah bank butir. Saat suatu test dibuat, seseorang memilih-milih arsip kartu untuk menentukan butir-butir yang tepat untuk test tersebut. Informasi butir yang dapat disimpan dalam suatu kartu adalah sebagai berikut:

(1) Informasi Identifikasi

- Nomer identifikasi butir
- Tahun kurikulum
- Level sekolah
- Level kelas
- Bidang subyek
- Program studi kuartal tahun sekolah
- Tujuan pembelajaran
- Topik
- Subtopik
- Indikator
- Kode uji coba
- Nomor pertanyaan pada uji coba
- Tanggal diuji-cobakan

(2) Informasi dari ITEMAN

- Point-biserial, untuk jawaban yang benar dan setiap pengecoh
- Proporsi kesahan, untuk jawaban yang benar dan setiap pengecoh
- Respon yang dijadikan kunci (**keyed respon**)
- Informasi dari BIGSTEPS (untuk bank butir calibrated)
- Point-biserial, untuk jawaban yang benar

- Index kesulitan butir (**calibration**)
- Standard error dari index kesulitan butir
- Statistik pelengkap standar
- Besar sampel

(3) Informasi Dasar Lain

- Text butir
- Sumber butir

Tentu, jika ada informasi butir lain yang dianggap penting, informasi tersebut bisa dimasukkan ke dalam kartu tersebut.

11) Pembuatan Tes Dengan Menggunakan Suatu Bank Butir

Setelah kita membuat bank butir, dan telah menyimpannya dalam kartu atau komputer, kita bisa menggunakan bank-bank butir ini dalam suatu test dan ujian. Hal ini, sesuai dengan tujuan pengembangan bank butir. Diharapkan, suatu bank butir memberi kita sekumpulan butir-butir yang bisa diambil dengan kualitas tinggi. Yang harus kita lakukan hanyalah memilih butir yang sesuai dari bank butir ini untuk keperluan ujian.

Ada beberapa program perangkat lunak yang bisa membuat pemilihan butir ini dilakukan secara otomatis berdasarkan pada kriteria yang diberikan pengguna. Namun, sampai saat ini pemilihan butir di Pusisjian dilakukan secara manual, demikian juga pada instansi-instansi lain dilingkungan Diknas. Oleh karena itu pembahasan kita pada kesempatan ini hanya dibatasi pada pemilihan butir dari bank butir secara manual butir. Di bawah ini adalah beberapa prinsip untuk pelaksanaan pemilihan butir secara manual untuk suatu test:

- a) Pilih butir yang memenuhi spesifikasi test untuk test yang dibutuhkan.
- b) Dalam memilih butir untuk mewakili suatu subtopik atau indikator tertentu, keluarkan semua butir yang memenuhi spesifikasi yang relevan, dan pilih yang paling cocok untuk test tersebut.
- c) Untuk bank butir **calibrated**, pusatkan tingkat kesulitan butir yang dimasukkan pada populasi target untuk test tersebut.
- d) Jika bentuk-bentuk pilihan ganda sedang dibuat dalam suatu subyek tertentu, cobalah untuk mendapatkan kesulitan butir rata-rata yang sama untuk bentuk test yang berbeda.

- e) Gunakan beragam tingkat kesulitan butir di dalam setiap bentuk test, bentangkan tingkat kesulitan butir antara butir yang paling mudah dengan butir yang paling sulit.
- f) Sebisa mungkin tidak mencoba menggunakan butir yang sama pada bentuk test yang berbeda untuk suatu subyek.
- g) Bersedialah untuk menggunakan butir-butir yang ditulis dan digunakan dalam tahun-tahun sebelumnya.

12) Memelihara Bank Butir dan Sistem Perbankan Butir

Setelah suatu sistem perbankan butir terbentuk, dan satu set bank butir telah dikembangkan untuk sistem ini, kita menginginkan agar sistem dan bank butir ini memberikan layanan yang berguna untuk periode waktu yang lama, yang memungkinkan kita untuk melaksanakan program pengujian yang berkualitas secara berkelanjutan. Agar bank butir dan sistem bank butir bisa tetap berfungsi baik, mereka harus dipelihara secara berkala. Maka kegiatan pemeliharaan ini harus dianggap sebagai komponen mendasar dari sistem perbankan butir itu sendiri.

Pertama, kita mempertimbangkan beberapa prinsip untuk pemeliharaan suatu bank butir; antara lain :

- a) Terus menambahkan butir-butir test baru ke dalam bank butir.
- b) Menghilangkan butir yang ketiggalan zaman dari bank butir.
- c) Tetap up-to-date-kan informasi mengenai penggunaan butir.
- d) Untuk bank butir yang sudah terkalibrasi, secara berkala skala ulang index-index kesulitan butir sehingga rata-ratanya adalah nol.
- e) Untuk bank butir yang sudah terkalibrasi, kalibrasi ulang butir dari waktu ke waktu, jika ada kesempatan yang muncul.

Sementara itu, prinsip-prinsip pemeliharaan suatu sistem perbankan butir adalah sebagai berikut.

- a) Pelihara semua bank butir dalam sistem perbankan butir terkait dengan cara di atas.
- b) Siapkan suatu jadwal aktivitas standar untuk dilaksanakan dalam menggunakan sistem perbankan butir, dan berpeganglah pada jadwal ini.
- c) Tetapkan tanggung jawab khusus bagi personel-personel yang terlibat dalam sistem perbankan butir tersebut, dan pastikan bahwa personel-personel ini melaksanakan tanggung jawab mereka tepat pada waktunya.

- d) Tetapkan dukungan keuangan untuk sistem perbankan butir tersebut sebagai butir anggaran reguler.
- e) Adakan rapat reguler para personel yang terlibat untuk membahas kemajuan dan masalah yang terkait dengan sistem perbankan butir tersebut.

Uraian di atas memberikan gambaran bahwa dengan adanya bank butir ini akan membuat tugas pembuatan test menjadi lebih mudah, dan akan bekerja untuk memastikan bahwa produk berkualitas tinggi akan dimunculkan. Terakhir, aktivitas pemeliharaan yang dilaksanakan secara berkala pada bank-bank butir ini akan menjamin keberfungsian mereka secara berkesinambungan dalam jangka panjang.

c. Karakteristik Butir Soal EBTANAS

Berbeda dengan tahun-tahun sebelumnya yang sebagian soal EBTANAS terdiri dari soal obyektif (khususnya pilihan ganda) dan sebagian lagi tes uraian, mulai tahun 1999 semua soal EBTANAS berbentuk tes obyektif.. Hal ini dilakukan karena obyektivitas soal tes bentuk uraian sangat diragukan. Bahkan, tes uraian ini merupakan suatu kesempatan bagi guru untuk membantu atau mengkatrol nilai siswa-siswinya.

Selanjutnya, untuk mendapatkan butir-butir soal yang berkualitas dilakukan analisis butir. Untuk soal-soal EBTANAS tingkat SD, hanya dianalisis secara teoritik atau telaah butir. Untuk soal-soal EBTANAS tingkat SLTA, sebagian diambil dari bank soal yang berarti sudah dianalisis secara teoritik dan analisis secara empirik, dan sebagian lagi merupakan hasil karya guru yang berarti hanya dianalisis secara teoritik saja. Sedangkan untuk soal-soal EBTANAS tingkat SMP hampir semuanya sudah dilakukan analisis soal secara teoritik dan analisis soal secara empirik.

Analisis teoritik atau telaah butir dilakukan dengan maksud menilai kualitas butir soal secara teoritis dipandang dari sudut : materi, konstruksi, dan bahasa. Pelaksana telaah butir ini adalah para guru yang dianggap menguasai cara menyusun soal tes yang baik dan menguasai materi yang diujikan. Sekali lagi, soal-soal EBTANAS yang hanya dikenai analisis teoritik ini adalah soal -soal yang tidak diambil dari bank soal.

Untuk soal EBTANAS yang diambil dari bank soal, proses analisisnya agak panjang. Selain dianalisis secara teoritik soal ini juga dianalisis secara empirik. Butir soal yang telah lolos dari telaah butir kemudian diuji-cobakan dan dianalisis dengan ITEMAN yang selanjutnya ditelaah lagi. Butir soal yang lolos dari telaah kedua, selanjutnya

dianalisis dengan menggunakan BIGSTEPS. Butir soal yang cocok masuk ke bank soal, dan yang tidak cocok direvisi lagi. Dengan demikian dalam analisis butir secara empirik ini digunakan dua metode, yaitu klasik dan metode Rasch.

Dalam analisis butir dengan menggunakan metode klasik, tiap-tiap butir soal dihitung daya beda, tingkat kesulitan, dan kefungsi-an pengecoh. Selain itu juga dapat dihitung reliabilitas soal. Sedangkan dengan menggunakan metode Rasch dihitung skala logit, Standar Error, Skala logit dan Fit-nya. Karakteristik tiap-tiap butir soal dituliskan dalam kartu soal yang juga memuat kode soal, tahun, nomor urut, tahun kurikulum, mata pelajaran, kelas/cawu, program studi, nomor tujuan pembelajaran, pokok bahasan, sub pokok bahasan, dan indikator.

Untuk soal EBTANAS tingkat SD, terdiri dari 24 % materi klas 4, 24 % materi klas 5, dan 52 % materi klas 6. Untuk soal EBTANAS tingkat SLTP, terdiri dari 24 % klas 1, 32 % materi klas 2, dan 44 % materi klas 3. Sedangkan untuk soal EBTANAS tingkat SMU terdiri dari 30 % klas 1, 34 % materi klas 2, dan 46 % materi klas 3.

d. Pelaksanaan Ujian dan Pemeriksaan Hasil Ujian

Ujian EBTANAS dilaksanakan di sekolah yang ditunjuk oleh Kanwil Depdiknas, jadi tidak semua sekolah ditunjuk sebagai tempat ujian EBTANAS. Sedangkan pengawas ujian di kelas/ruang adalah guru yang diusulkan oleh kepala sekolah kepada ketua penyelenggara EBTANAS tingkat sub-rayon. Untuk menjaga obyektivitas, pengawas ujian dilakukan secara silang. Guru tidak diperbolehkan mengawasi siswa-siswinya sendiri atau siswa yang diajar setiap hari.

Untuk pengamanan soal ujian EBTANAS, menurut hasil penelitian Mardapi dan Kartowagiran (1999), telah dilakukan beberapa tindakan, yaitu:

- 1) Pada saat penyusunan, pengembangan dan penyuntingan soal di wilayah, tim penyusun disumpah untuk tidak membocorkan rahasia negara dan dibimbing langsung oleh Tim Pusiasjian Jakarta, serta dijaga dari fihak kepolisian.
- 2) Pada saat penggandaan naskah di wilayah, diawasi dari Kanwil Dekdiknas, Dinas Pdan K Dati II, Kanwil Depag, Petugas Kepolisian, dan Tim Koreksi dari propinsi.
- 3) Pada saat pengiriman soal dari Kanwil (tempat penggandaan) ke rayon diantar oleh petugas dari Kanwil Depdiknas, Dinas P dan K Dati II, dan petugas dari kepolisian.
- 4) Dari sub rayon, soal diambil oleh petugas dari sekolah pada hari ujian

5) Pada saat ujian berlangsung, tempat duduk peserta ujian berjarak kurang lebih 0,6 meter, diawasi oleh guru dari sekolah lainnya.

Pengamanan soal tidak hanya dilakukan di tingkat Kanwil, Rayon ataupun sub rayon, tetapi juga dilakukan di sekolah-sekolah tempat ujian. Hal-hal yang dilakukan adalah mengambil soal pada hari ujian, pengambilan soal dari sub rayon didampingi oleh petugas keamanan atau polisi setempat. Setelah sampai di tempat ujian soal disimpan oleh Kepala Sekolah sampai waktu ujian tiba. Dengan demikian sebagai penanggung jawab keamanan soal di sekolah adalah Kepala Sekolah, termasuk di dalamnya saat ujian berlangsung.

Lembar jawaban siswa (untuk tingkat SLTP dan SMU di sebagian besar propinsi telah menggunakan lembar jawab komputer) dimasukkan ke dalam amplop dan ditutup rapat di ruang/kelas tempat ujian. Selanjutnya, bersama dengan buku soal ujian, lembar jawaban siswa dikirim ke sub rayon. Di sub rayon, lembar jawaban siswa disusun atau diurutkan kemudian dikirim ke rayon atau panitia tingkat kabupaten/kotamadya. Setelah lembar jawaban siswa dari seluruh sub rayon terkumpul selanjutnya dikirim ke Kanwil diserahkan ke panitia penyelenggara EBTANAS tingkat propinsi. Lembar jawaban siswa SLTP dan SMU yang menggunakan lembar jawab komputer dikoreksi di Kanwil dengan menggunakan komputer.

Sedangkan untuk tingkat SD dan SLTP, SLTA yang belum menggunakan lembar jawab komputer, sesudah digunakan, soal disimpan di sub rayon. Demikian pula lembar jawaban siswa yang belum dikoreksi juga disimpan di sub rayon. Pemindehan lembar jawaban siswa yang belum dikoreksi dari sekolah ke sub rayon dilakukan pada hari ujian juga. Selanjutnya , sesuai dengan jadwal yang telah ditentukan, lembar jawaban siswa dikoreksi di sub rayon secara manual. Dalam koreksi secara manual ini, setiap lembar jawaban dikoreksi oleh dua orang guru dan dilakukan secara silang. Nilai suatu mata pelajaran hasil EBTANAS adalah rerata dari hasil koreksi kedua korektor tersebut. Untuk soal yang berbentuk uraian, apabila ada perbedaan hasil koreksi antara korektor I dan korektor II lebih dari 20 % harus dilakukan koreksi ulang oleh korektor III. Nilai akhirnya adalah rata-rata nilai dari ketiga korektor tersebut. Sedangkan untuk soal pilihan ganda atau isian singkat, apabila terjadi perbedaan antara korektor I dan II lebih dari 1 butir soal dilakukan pemeriksaan ulang. Hasil koreksi manual ini selanjutnya dilaporkan

ke Kanwil Depdiknas. Di Kanwil, baik hasil koreksi dengan komputer maupun dengan manual, hasil ujian EBTANAS ini diperiksa ulang. Setelah dianggap selesai, nilai EBTANAS ini selanjutnya dikembalikan ke sekolah-sekolah untuk diumumkan.

Hasil EBTANAS ini biasanya disebut dengan Nilai Ebtanas Murni (NEM). Sampai saat ini NEM digunakan sebagai alat seleksi masuk jenjang ke sekolah yang lebih tinggi. Bahkan, masyarakat agak berlebihan menghargai NEM ini. Hal ini terbukti dari adanya kenyataan bahwa sebagian besar masyarakat menganggap suatu sekolah dikatakan berkualitas apabila sekolah itu mampu meluluskan siswanya dengan NEM tinggi. Mestinya NEM hanya salah satu indikator berkualitas atau tidak berkualitasnya suatu sekolah.

2. Tes of Spoken English (TSE) yang Dikelola ETS

a. Pengertian

Hampir 49 tahun (1951 - 2000) Educational Testing Service (ETS) telah terjun ke dunia internasional dalam bidang Penelitian dan Penilaian Pendidikan (Toefel Products & services Catalog, 1997). Dalam kurun itu ETS telah mengeluarkan beberapa tes bertaraf internasional, yaitu: Test of English as a Foreign Language (TOEFL), Test of Written English (TWE), Secondary Level English Proficiency (SLEP) test, Speaking Proficiency English Assesment Kit (SPEAK), Institutional Testing Program (ITP), dan Test of Spoken English (TSE). Jadi, *Tes of Spoken English (TSE)* adalah salah satu jenis tes yang dikeluarkan oleh *Educational Testing Service (ETS)*.

Tujuan TSE adalah untuk mengukur kemampuan orang yang bahasa-ibu-nya bukan bahasa Inggris (*nonnative speakers of English*) untuk berkomunikasi secara lisan dalam bahasa Inggris. TSE disebar-luaskan dalam suatu format yang semi langsung yang selalu dijaga reliabilitas dan validitasnya sementara pengontrolan variabel subyektifnya dibantu dengan wawancara langsung. Oleh karena TSE merupakan suatu tes tentang kemampuan berbahasa lisan (oral) secara umum maka tes jenis ini cocok untuk hampir semua peserta tes tanpa memperhatikan: bahasa-ibu, tipe latihan, atau bidang pekerjaan yang ditekuni peserta tes.

TSE mempunyai nilai guna yang sangat tinggi karena hasil tes ini dapat menggam-barkan kemampuan seseorang dalam berbahasa lisan baik akademik maupun lingkungan profesional (Score User's manual, 1995). Skor TSE digunakan sebagai alat

seleksi men-jadi asisten guru internasional (International Teacher Assistants = ITAs) untuk institusi pendidikan tinggi pada banyak negara di Amerika Utara. Skor TSE juga digunakan untuk seleksi masuk menjadi tenaga medis, seperti dokter, perawat, farmasi, dan dokter hewan.

Skor TSE seharusnya tidak diinterpretasikan sebagai prediktor keberhasilan dalam bidang akademik dan profesional, namun hanya sebagai indikator kemampuan seseorang yang bahasa-ibu-nya bukan bahasa Inggris dalam berkomunikasi dalam bahasa Inggris. Skor TSE hendaknya digunakan sebagai informasi lain tentang calon bila akan membuat keputusan mengenai kemampuan calon untuk tampil dalam situasi akademik maupun profesional.

b. Perkembangan TSE

TSE awal berkembang pada akhir tahun 1970-an seiring adanya fakta bahwa lembaga-lembaga akademik sering kali membutuhkan suatu alat pengukur yang akurat untuk kemampuan berbicara untuk kebutuhan pembuatan keputusan perekrutan pegawai dan pemilihan yang teliti. Pada saat itu penekanan ditempatkan pada bidang ilmu linguistik, pengajaran bahasa dan pengetesan bahasa dalam keakuratan pelafalan, tata bahasa, dan kefasihan. Tes tersebut dirancang untuk mengukur unsur-unsur kebahasaan tersebut dan untuk mengevaluasi kemampuan seorang pembicara untuk menyampaikan informasi dengan bisa dimengerti kepada pendengarnya. Skor-skor tes didapatkan dari pelafalan, tata bahasa, kefasihan, dan kedapat-dimengertian secara keseluruhan.

Pada tahun 1978 TOEFL Research Committee dan TOEFL Policy Council mensponsori suatu study yang berjudul "An Exploration of Speaking Proficiency Measures in the TOEFL Context" (Clark dan Swinton, 1979). Laporan dari studi ini merinci prosedur dan alasan pengukuran yang digunakan untuk mengembangkan TSE, dan juga sebagai dasar pemilihan format dan jenis pertanyaan tertentu yang dimasukkan dalam bentuk test yang awal.

Satu pertimbangan utama dalam pengembangan suatu alat pengukur kemampuan berbicara adalah bahwa pengukur tersebut harus sesuai dengan standar pelaksanaan (administrasi) pada pusat-pusat test TOEFL. Faktor ketiga ini segera menghilangkan praktik wawancara langsung yang saling berhadapan. Pemberian pelatihan yang

diperlukan dalam teknik wawancara dengan pertimbangan penggunaannya di seluruh dunia dianggap tidak praktis.

Faktor lain yang terkait selama perkembangan TSE awal adalah kandungan linguistiknya. Karena tes ini akan diterapkan di banyak negara, tes ini harus sesuai untuk semua peserta tes tanpa melihat bahasa atau budaya aslinya.

Faktor ketiga dalam pertimbangan rancangan tes adalah kebutuhan untuk mendapatkan bukti-bukti kemampuan berbicara secara umum bukannya kemampuan dalam dalam suatu situasi penggunaan bahasa. Karena tes tersebut akan digunakan untuk memperkirakan kemampuan berbicara peserta ujian dalam beragam konteks, tes ini tidak dapat menggunakan format item atau pertanyaan tersendiri yang akan membutuhkan pengetahuan yang luas dalam suatu subyek atau konteks pekerjaan tertentu.

Dua bentuk pengembangan TSE telah diterapkan pada 155 peserta ujian, yang juga mengambil TOEFL dan berpartisipasi dalam suatu wawancara kemampuan lisan sesuai model yang diterapkan pada Foreign Service Institute (FSI). Item-item spesifik yang dimasukkan ke dalam bentuk prototipe tersebut dipilih dengan tujuan mempertahankan kemungkinan korelasi tertinggi dengan rating FSi dan kemungkinan korelasi terendah dengan skor TOEFL untuk memaksimalkan kebergunaan tes berbicara.

Validasi TSI didukung oleh riset yang menunjukkan hubungan antara nilai kedapat-dimengertian TSE dan level kemampuan lisan FSi, interkorelasi antara keempat skor TSE, dan korelasi skor TSE instruktur universitas dengan penilaian siswa terhadap skill bahasa instruktur (Clark dan Swinton, 1980).

Setelah pengenalan tes TSE untuk penggunaan oleh lembaga-lembaga akademik pada tahun 1981, riset tambahan (Powers dan Stansfield, 1983) menvalidasi skor-skor TSE untuk pemilihan dan serifikasi dalam profesi yang terkait dengan kesehatan (misal, obat-obatan, keprawatan, farmasi, dan obat-obatan ternak).

c. Revisi Test TSE

Sejak pengenalan TSE awal pada tahun 1981, teori dan praktik pengajaran dan pengetesan bahasa berkembang untuk menempatkan penekanan pada kemampuan bahasa komunikatif lisan. Pendekatan kontemporer ini memasukan keakuratan lingistik hanya sebagai salah satu dari beberapa aspek kompetensi bahasa yang terkait dengan

keefektifan komunikasi lisan. Untuk alasan ini, TSE telah direvisi agar bisa lebih baik dalam mencerminkan pandangan penilaian dan kemampuan bahasa yang sekarang digunakan.

Pada April 1992 TOEFL Policy Council menyetujui rekomendasi TOEFL Committee of examiners untuk merevisi TSE dan untuk membentuk suatu komite TSE terpisah untuk mengawasi usaha revisi terkait. Anggota-anggota komite TSE dipilih oleh TOEFL Policy Council Executive Committee. Anggota TSE mencakup para spesialis dalam bidang linguistik terapan serta pengajaran dan pengetesan bahasa Inggris lisan, dan juga perwakilan para pengguna skor TSE. Komite TSE menyetujui spesifikasi tes dan skala nilai, meninjau pertanyaan tes dan performa item, menawarkan panduan dalam pelatihan penilai (*rater*) dan penggunaan skor, dan memberi saran-saran untuk riset selanjutnya.

Para anggota komite TSE dirotasi secara berkala untuk memastikan keberlanjutan pengenalan ide-ide dan perspektif yang baru dalam kaitannya dengan penilaian kemampuan bahasa lisan. Lampiran A mendaftar anggota-anggota komite TSE pada saat ini dan anggota yang lama.

Proyek revisi TSE ini, yang dimulai pada tahun 1992, merupakan suatu usaha bersama antara staf ETS dan komite TSE. Proyek yang terkonsentrasi selama 3 tahun ini memerlukan penguatan basis teori yang mendasari tes dan spesifikasi tes serta revisi skala peratingan. Riset pengembangan mencakup pengetesan awal yang ekstensif untuk item tes dan bahan peratingan, prototipe studi riset skala besar, dan serangkaian studi untuk memvalidasi tes yang direvisi dan sistem penskoran. Publikasi program terkait mengalami revisi yang ekstensif, dan suatu TSE Standars-Setting Kit dibuat untuk membantu para pengguna dalam menetapkan skor passing untuk TSE yang direvisi.

Pelatihan *rater* yang ekstensif juga dilaksanakan untuk men-set standar peratingan dan memastikan penerapan yang tepat untuk sistem penskoran yang direvisi. Pada awal proyek revisi TSE, disetujui bahwa tujuan tes tetap tidak berubah. Yaitu, TSE akan terus menjadi suatu tes kemampuan berbicara secara umum yang dirancang untuk mengevaluasi kemampuan bahasa lisan dari penutur Bahasa Inggris bukan asli yang ada pada atau melebihi level pendidikan "postsecondary". TSE akan terus menjadi hal yang bermanfaat bagi audiens utama dari TSE awal (yaitu, pihak yang mengevaluasi calon ITA

dan personel untuk profesi yang terkait dengan kesehatan). Dalam hal ini, TSE dirancang sebagai suatu alat pengukur kemampuan peserta ujian untuk berhasil berkomunikasi dalam bahasa Inggris di lingkungan akademik atau profesi.

Juga dinyatakan bahwa TSE akan terus menjadi tes berbicara semi langsung yang dilaksanakan melalui peralatan audio-recording yang menggunakan kaset rekaman dan buku tes tercetak, dan bahwa respon terekam peserta tes, akan diskor setidaknya oleh dua rater terlatih. Tes awal (pilot) untuk tiap bentuk tes memungkinkan ETS untuk memonitor performa dari semua pertanyaan tes.

1) Konstruksi Tes

Komite TSE menggunakan suatu karya ilmiahnya Douglas dan Smith (1994) untuk memberikan suatu tinjauan catatan riset, menguraikan asumsi teoritis mengenai kemampuan berbicara, dan berfungsi sebagai panduan untuk revisi tes. Karya ilmiah ini, "Theoretical Underpinnings of the Test of Spoken English Revision Project", menjelaskan model-model penggunaan dan kompetensi bahasa, dengan menekankan pada bagaimana model-model tersebut mungkin bisa membantu rancangan dan penskoran tes. Karya ini juga menyatakan batasan-batasan tes secara audio dibandingkan suatu wawancara langsung.

Seperti tersirat dari tulisan teori tersebut, construct yang mendasari tes yang direvisi adalah kemampuan bahasa komunikatif. Tes TSE direvisi berdasar premis bahwa bahasa merupakan kendaraan yang dinamis untuk komunikasi, dengan diarahkan oleh kompetensi dasar yang berinteraksi dalam beragam cara agar suatu komunikasi yang efektif bisa berlangsung. Untuk tujuan TSE, kemampuan bahasa komunikatif ini dibuat agar mencakup kompetensi strategis dan kompetensi bahasa, kompetensi bahasa ini terdiri dari kompetensi linguistik, kompetensi wacana, kompetensi fungsional, dan kompetensi sociolinguistik.

Hal yang penting untuk rancangan test adalah gagasan bahwa kompetensi-kompetensi ini tercakup dalam aktivitas komunikasi yang sukses. Penggunaan bahasa untuk tujuan atau fungsi yang disengaja (misalnya untuk memaafkan, untuk mengeluh) adalah hal yang penting untuk komunikasi yang efektif. Oleh karena itu setiap item test terdiri dari tugas bahasa yang dirancang untuk mendapatkan fungsi khusus dalam konteks atau situasi tertentu.

Dalam kerangka kerja ini, keragaman tugas dan fungsi-fungsi bahasa dibuat agar memberikan dasar dari test revisi. Sistem penilaian juga dirancang untuk memberi ringkasan holistik pada kemampuan bahasa lisan pada kompetensi-kompetensi komunikasi yang sedang dinilai.

2) Validitas test

Rangkaian dari aktivitas validisasi telah dilakukan selama revisi TSE untuk mengevaluasi kelayakan rancangan test dan untuk memberi bukti kegunaan nilai-nilai TSE. Upaya ini dilaksanakan dalam sudut pandang yang berorientasi pada proses. Yaitu akumulasi data validitas yang digunakan untuk membantu revisi test, membuat modifikasi seperti yang disarankan, dan mengkonfirmasi kelayakan dari rancangan test dan skala penilaian.

Validitas yang mengacu pada sejauh mana suatu tes benar-benar mengukur apa yang seharusnya diukur. Walaupun banyak prosedur ada ditemukan untuk penentuan validitas, tapi tidak ada satupun indikator atau index standar untuk validitas. Sejauh mana suatu tes bisa dievaluasi sebagai suatu alat pengukur yang valid ditentukan dengan menilai semua bukti yang tersedia. Keunggulan dan batasan tes dan juga kecocokannya untuk penggunaan tertentu dan populasi peserta tes harus dipertimbangkan.

Riset validitas construct dimulai dalam tulisan teori yang digunakan oleh komite TSE (Douglas dan Smith, 1994). Dokumen ini membahas sifat dinamis dari construct kemampuan bahasa lisan di bidang penilaian bahasa dan menunjukkan arah kepada suatu basis konseptual untuk tes revisi. Sebagai hasil dari tulisan tersebut dan diskusi para ahli di bidang tersebut, construct dasar yang mendasari tes didefinisikan sebagai kemampuan bahasa komunikatif. Konsep teoritis ini dioperasionalkan pada spesifikasi tes awal.

Untuk mengevaluasi validitas rancangan tes, Hudson (1994) meninjau tingkat kesamaan antara basis teori tes dan spesifikasi tes. Analisa ini menunjukkan adanya tingkat kesesuaian yang secara umum yang tinggi. Spesifikasi tes kemudian direvisi dengan adanya tinjauan ini.

Dengan cara yang sama, tes prototipe juga diteliti oleh staf ETS untuk mencari tingkat kesesuaiannya dengan spesifikasi tes. Tinjauan ini juga mengarah pada revisi sederhana dalam spesifikasi tes dan garis panduan penulisan item untuk memberikan tingkat kesesuaian yang tinggi antara teori, spesifikasi, dan bentuk tes.

Sebagai suatu sarana validasi kandungan tes, suatu analisa wacana dari tuturan penutur asli dan non asli yang diambil oleh test prototipe telah dilaksanakan (Lazarton dan Wagner, 1994). Analisa ini menunjukkan bahwa fungsi bahasa yang disengaja yang reliable dan konsisten diambil dari penutur asli dan non asli, yang semuanya itu melaksanakan tipe aktivitas bicara yang sama.

Skala peratingan tes dan bidang skor divelidasi melalui satu proses yang lain. Staf peratingan ETS menuliskan deskripsi-deskripsi bahasa yang diambil dalam sampel tuturan yang dibandingkan dengan skala rating dan bidang skor yang diterapkan untuk sampel tersebut. Hal ini dilakukan untuk menentukan tingkat kesesuaian antara tuturan yang diambil dengan sistem penskoran. Hasilnya ternyata menguatkan validitas sistem peratingan.

Validitas concurrent dari TSE revisi diselidiki dalam studi riset skala besar oleh Henning, Schedl, dan Suomi (1995). Samael untuk studi ini terdiri dari subyek-subyek yang mewakili populasi utama peserta tes TSE: calon asisten pengajar an universitas (N = 184) dan calon profesional medis berlisensi (N = 158).

Calon-calon asisten mengajar mewakili bidang sains, teknik, ilmu komputer, dan ilmu ekonomi. Calon profesional medis berlisensi mencakup lulusan medis asing yang sedang mencari lisensi untuk melakukan praktik sebagai dokter umum, perawat, dokter hewan, atau ahli farmasi di Amerika Serikat. Subyek-subyek di kedua kelompok mewakili lebih dari 20 bahasa ibu.

Instrumen yang digunakan dalam studi tersebut mencakup suatu versi awal TSE, satu versi prototipe revisi TSE dengan 15 item, dan suatu wawancara kemampuan bahasa (LPI) lisan. Versi awal dan prototipe revisi diterapkan di bawah persyaratan standar TSE.

Studi ini menggunakan dua tipe rater: 16 rater yang kurang secara linguistik dan tidak terlatih, dan 40 rater ahli yang terlatih. Rater yang kurang, 8 dari populasi siswa dan 8 dari calon populasi pasien medis, dipilih karena mereka mewakili kelompok yang paling mungkin terpengaruhi oleh kemampuan bertutur bahasa Inggris dari kandidat penutur non-asli yang membutuhkan skor kelulusan TSE. Rater-rater ini sengaja dipilih karena mereka hanya punya sedikit pengalaman dalam berinteraksi dengan penutur non asli bahasa Inggris, dan hanya menilai respon-respon untuk prototipe revisi. Para rater kurang diminta untuk menilai keefektifan komunikasi dari respon prototipe TSE revisi

dari 39 subyek sebagai bagian validasi metode penskoran revisi. Para rater terlatih menskor performa peserta tes pada TSE awal berdasar pada skala peratingan awal dan menilai performa pada prototipe revisi TSE berdasar pada skala peratingan. (Skala peratingan yang digunakan dalam studi ini untuk menskor TSE revisi adalah serupa walaupun tidak sama dengan skala peratingan akhir.)

Penggunaan rater kurang (naif) dalam studi ini berfungsi untuk memberikan bukti validitas konstruk tambahan untuk inferensi yang akan diambil dari skor tes. Para rater naif yang tak terlatih mampu menentukan dan membedakan beragam level kemampuan bahasa komunikatif dari sampel performa tuturan yang diambil oleh test prototipe. Hasil-hasil ini juga memberikan *content validity* untuk bidang skala peratingan dan interpretasi skor.

Rata-rata dan standar deviasi dihitung untuk skor yang diberikan oleh perater terlatih. Dalam studi awal ini, rata-rata skor pada prototipe TSE revisi adalah 50,27 dan standar deviasi adalah 8,66. Perbandingan yang dibuat dari performa subyek pada TSE awal dan prototipe tes revisi menunjukkan suatu korelasi antara skor untyuk kedua versi adalah 0,83.

Sebagai bagian dari studi riset, suatu subsampel dari 39 peserta tes diberlakukan suatu wawancara kemampuan bahasa (LIP) lisan formal yang diakui oleh American Council on the Teaching of Foreign Languages (ACTFL), Foreign Service Institute (FSI), dan Interagency Language Roundtable (ILR). Korelasi antara skor pada LPI dan prototipe TSE ternyata 0,82, yang memberikan bukti lanjutan dari validitas concurrent untuk revisi tes.

3) Reliabilitas tes

Reliabilitas tes didefinisikan sebagai sejauh mana suatu tes menghasilkan hasil pengukuran yang konsisten. Reliabilitas interrater adalah suatu alat pengukur kekonsistenan skor di antara para rater. Dalam studi prototipe (Henning dkk, 1995), reliabilitas interrater adalah 0,82 untuk seorang rater terlatih. Koefisien ini mewakili korelasi sebenarnya antara skor-skor yang diberikan oleh dua rater yang menskor rekaman peserta tes yang sama. Jika skor peserta tes didasarkan pada rata-rata kedua penilaian, reliabilitas interrater untuk peratingan prototipe adalah 0,90, yang menunjukkan suatu tingkat konsistensi yang tinggi. jika skor akhir yang diberikan oleh kedua rater

berbeda jauh, seorang rater ketiga dibutuhkan untuk menskor tes tersebut. Perkiraan konsistensi reliabilitas internal (koefisien alpha) adalah 0,96.

d. Kandungan dan Format Program TSE

TSE terdiri dari 12 item, yang tiap itemnya mengharuskan para peserta tes untuk melakukan suatu aksi bertutur tertentu. Contohnya, yang disebut juga fungsi bahasa, mencakup narasi, pemebrian saran, membujuk, dan memberi dan mendukung suatu pendapat. Tes tersebut disampaikan melalui peralatan audio-recording dan suatu buku tes. Seorang pewawancara pada rekaman tes memandu peserta tes untuk melaksanakan tes; peserta tes merespon ke dalam suatu mikropon, dan respon-respon tersebut direkam pada kaset jawaban yang terpisah.

Waktu yang dialokasikan untuk tiap respon berkisar dari 30 sampai 90 detik. Semua pertanyaan yng ditanyakan oelh pewawancara,dan juga waktu respon dicetak dalam buku tes. pertanyaan-pertanyaan pada tes tersebut bersifat umum dan dirancang untuk memberi informasi pada rater mengenai kemampuan bahsa komunikasi lisan peserta tes.

Pada awal tes, pewawancara pada kaset tes memberikan beberapa pertanyaan umum yang berfungsi sebagai "pemanasan" untuk membantu peserta ujian unuk menjadi terbiasa untuk berbicara di kaset dan memungkinkan penyesuaian peralatan audio seperlunya. Kemudian, peserta tes diberi waktu 30 detik untuk mempelajari suatu peta dan lalu diberi pertanyaan mengenai peta tersebut. Kemudian, peserta tes diminta mengamati serangkaian gambar dan menceritakan kembali cerita yang ditunjukkan gambar-gambar tersebut. Lalu peserta tes diminta membahas informasi yang diberikan pada suatu grafik sederhana. terakhir, peserta tes diminta untuk menyajikan informasi dari suatu jadwal revisi dan menunjukkan revisi yang sudah dilakukan.

1) Pendaftaran Tes

Tanggal pendaftaran tes dipublikasikan di "Bulletin of Information for TOEFL, TWE, and TSE". Salinan buletin ini didistribusikan kepada pusat-pusat tes TSE dan TOEFL, ke kedutaan-kedutaan besar USA, pusat-pusat kerjasama bilateral, akademi bahasa, serta agensi-agensi tambahan dan perseorangan yang menunjukkan minat pada TSE. Seringkali lembaga-lembaga atau departmen dan employer yang membutuhkan nilai TSE dari pelamar mengikutkan salinan buletin ini saat merespon pada permohonan

kerja dari penutur bahasa Inggris non asli. Buletin ini mencakup satu formulir pendaftaran, deskripsi umum tentang tes, petunjuk tes, dan pertanyaan latihan. Kandidat TSe harus mengisi formulir pendaftaran dan menyerahkannya pada TOEFL/TSE Service bersama dengan pembayaran tes yang diperlakukan. Buletin ini juga bisa didapat dari TOEFL/TSE Service, PO Box 6151, Princeton, NJ 08541-6151, USA. (lihat formulir pemesanan pada hal. 23 untuk memesan lebih dari lima eksemplar buletin.)

TSE dilaksanakan 12 kali setahun di pusat-pusat tes di seluruh dunia dengan prosedur pengetesan yang dikontrol secara ketat. Waktu pengetesan sebenarnya adalah sekitar 20 menit. Tes ini bisa dilaksanakan pada individual dengan cassette tape recorder atau pada satu kelompok dengan menggunakan fasilitas perekaman ganda misalnya laboratorium bahasa.

Karena skor-skor peserta tes baru bisa setara hanya jika prosedur yang sama diikuti pada semua pelaksanaan tes, TSE Program Office memberikan garis panduan rinci untuk para pengawas pusat tes untuk menjamin pelaksanaan tes yang seragam. TSE Supervisor's Manual dikirimkan bersama dengan bahan tes pada pengawas tes jauh sebelum pelaksanaan tes. Publikasi ini menjelaskan tata cara yang diperlukan untuk mempersiapkan pelaksanaan tes, membahas jenis peralatan yang diperlukan, dan memberi instruksi rinci untuk pelaksanaan tes yang sebenarnya.

Para peserta ujian tidak diizinkan untuk membawa kertas, pena atau pensil, kamus, atau alat foto atau perekam pribadi ke dalam ruangan pengetesan, dan mereka tidak boleh memakai gelang, kalung atau perhiasana lain yang bisa membuat bunyi yang mengganggu saat tes sedang berlangsung.

Pada awal tes, sebelum permulaan tes sebenarnya, para peserta tes diberi buku tes yang tersegel. Saat tes dimulai, peserta tes menyimak suatu rekaman yang berisi petunjuk umum dan pertanyaan tes. Tape recorder yang dipakai merekam respon peserta tes tidak boleh dimatikan kapan pun selama tes berlangsung kecuali terdapat situasi tak biasa yang terkait dengan pelaksanaan tes teridentifikasi oleh pelaksana tes.

2) Layanan bagi individu yang cacat

TSE Program Office, dalam responnya pada permintaan dari individu yang cacat, akan membuat persiapan khusus dengan para pengawas pusat-pusat tes, di mana kondisi lokal memungkinkan, untuk melaksanakan TSE di bawah kondisi non standar. Dalam

pelaksanaan tes non standar, dibuat akomodasi untuk peserta tes dengan cacat pandangan, pendengaran, atau fisik. Dalam kasus cacat mata, hanya pertanyaan-pertanyaan tertentu yang bisa diberikan. Pertanyaan-pertanyaan yang mengharuskan peserta tes menjelaskan hal-hal visual di dalam buku tes tidak bisa dilaksanakan, sehingga tidak diperhitungkan dalam menghitung skor tes.

3) Langkah-langkah untuk keamanan tes

Untuk melindungi validitas skor tes, TSE Program Office secara konstan meninjau dan menyempurnakan prosedur-prosedur yang dirancang untuk meningkatkan keamanan tes sebelum, selama, dan sesudah pelaksanaan tes. Dikarenakan pentingnya skor TSE untuk pelamar dan untuk institusi, tak dapat dihindarkan munculnya beberapa individu yang terlibat dalam praktik-praktik yang dirancang untuk meningkatkan skor terlapor mereka. Pemilihan pengawas yang hati-hati, rasio peserta terhadap pengawas yang rendah, dan prosedur pelaksanaan yang rinci dalam Supervisor's Manual semuanya itu dirancang untuk mencegah usaha-usaha penipuan, pencurian bahan tes, dan sejenisnya, sehingga melindungi integritas tes untuk semua peserta tes dan penerima skor.

Prosedur masuknya peserta yang ketat diberlakukan di semua pusat tes untuk mencegah usaha-usaha para peserta tes untuk memasukan orang lain dengan kemampuan lebih besar dalam bahasa Inggris berpura-pura menjadi mereka pada pelaksanaan tes. Untuk bisa masuk pusat tes, setiap peserta tes harus menunjukansuatu dokumen identifikasi resmi dengan foto yang bisa dikenali, misalnya passport sah.

Selain passport yang menjadi dokumen basis yang ditemui di semua pusat tes, dokumen berfoto khusus lain bisa diterima untuk individu yang mungkin tidak bisa mendapatkan passport atau orang yang engambil tes di negara mereka sendiri. Melalui kedubes-kedubes asing di USA dan pengawas TSE di negara-negara asing, TOEFL/TSE Service memverifikasi tipe-tipedokumen-dokumen identifikasi berfoto resmi yang digunakan di tiap negara, misalnya KTP, sertifikay pendaftaran, dan izin kerja. Informasi rinci mengenai persyaratan identifikasi dimasukan ke dalam bulletin terkait.

Catan pribadi yang berisi nama peserta tes, nomor pendaftaran, kode pusat tes, tanda tangan dan juga foto terbaru yang dengan jelas mengidentifikasikan peserta tes. Formulir ini dikumpulkan oleh pengawas pusat tes dari tiap peserta tes sebelum dia diizinkan masuk ke dalam ruang pengetesan. Selain verifikasi identitas foto peserta tes,

pengawas memverifikasi apakah nama pada dokumen identifikasi resmi sama dengan nama pada catatan foto file.

Para pengawas (supervisor) dan pengawas ruangan diinstruksikan untuk menerapkan kewaspadaan yang tinggi selama pelaksanaan tes untuk mencegah para peserta tes untuk memberi dan mendapatkan bantuan dalam bentuk apapun. Saat mengambil tes, para peserta tes tidak boleh menyediakan apapun di atas meja selain buku tes mereka, alat perekam dan tanda masuk. Mereka tidak diperbolehkan membuat catatan atau tanda apapun dalam buku tes mereka.

Jika pengawas merasa yakin bahwa seseorang telah memberi atau menerima bantuan kepada orang lain, peserta tes tersebut dikeluarkan dari ruang tes dan skornya tidak dilaporkan. Jika pengawas curiga seseorang mencontek, peserta ujian tersebut diberikan peringatan atas pelanggaran itu. Penjelasan kejadian yang terjadi dituliskan pada Irregularity Report dari Supervisor (dimasukan dalam Supervisor's Manual), yang dikembalikan ke ETS dengan rekaman peserta tes. Kasus kecurangan yang dicurigai dan/atau terbukti diselidiki oleh Test Security Office di ETS.

Untuk menjamin bahwa peserta tes tidak melihat bahan tes sebelumnya, maka secara berkala selalu dikembangkan bentuk-bentuk tes yang baru. Selain itu, untuk mencegah adanya pencurian bahan tes, banyak prosedur yang dirancang untuk keamanan distribusi dan penyerahan bahan-bahan ini. Kaset tes dan buku tes (yang disegel terpisah dan dikemas dalam kantung plastik bersegel) dikirimkan ke pusat-pusat tes dalam kotak bersegel dan para pengawas diharuskan menempatkannya di ruangan penyimpanan terkunci yang tidak boleh dimasuki pihak yang tidak berwenang. Pengawas diperintahkan untuk menghitung buku tes saat diterima, setelah peserta tes memulai tes, and pada akhir pelaksanaan tes. Tak seorangpun diizinkan untuk meninggalkan ruangan tes sampai semua buku tes dan kaset rekaman jawaban peserta tes telah dihitung.

Petunjuk jelas diberikan kepada para pengawas untuk mengembalikan bahan tes pada ETS, di mana semuanya dihitung saat diterima. ETS dan Tes Security Office menyelidiki semua kasus hilangnya bahan tes.

e. Sistem Penyekoran TSE

Kaset jawaban TSE dinilai oleh rater TSE terlatih yang adalah guru-guru dan spesialis berpengalaman di bidang bahasa Inggris atau bahasa Inggris sebagai bahasa

kedua. Para rater dilatih pada workshop pengkualifikasian yang dilakukan oleh staf ETS. Sebelum tiap sesi penskoran tes, para rater mereview kaset jawaban di beragam titik pada skala rating TSE untuk menjaga penskoran yang akurat. Rater mengalami pelatihan kembali jika perbedaan skor yang mencolok menunjukkan bahwa hal ini diperlukan.

Tiap kaset TSE dinilai secara terpisah oleh dua rater; tak seorang pun skor yang diberikan yang diberikan rater yang lain. Tiap rater mengevaluasi tiap respon item dan memberikan satu level skor dengan menggunakan descriptor keefektifan komunikasi yang dinyatakan dalam skala rating TSE (lihat lampiran B). Skor-skor peserta tes dihasilkan dari rata-rata gabungan peratingan item terpisah tersebut. Jika kedua penilaian itu tidak menunjukkan kesesuaian yang memadai, kaset tersebut dinilai oleh rater independen ketiga. Skor akhir untuk kaset rekaman yang membutuhkan penilaian ketiga didasarkan pada resolution dari perbedaan-perbedaan di antara ketiga skor.

Skor TSE terdiri dari satu nilai tunggal untuk kemampuan bahasa komunikatif, yang dilaporkan pada satu skala 20 sampai 60. Level nilai yang diberikan dirata-ratakan dari item dan rater, dan skor dilaporkan dalam penambahan lima nilai (yaitu, 20, 25, 30, 35, 40, 45, 50, 55, dan 60). Performa level skor dijelaskan di bawah ini.

Skala	Deskripsi
60	Komunikasi hampir selalu efektif
55	
50	Komunikasi secara umum efektif
45	
40	Komunikasi agak efektif
35	
30	Komunikasi secara umum tidak efektif
25	
20	Tidak ada komunikasi yang efektif

Jika respon terhadap salah satu item hilang, tidak ada skor tes yang dilaporkan.

Dua tipe catatan skor dikeluarkan untuk TSE: catatan skor peserta tes, yang dikirimkan langsung kepada peserta tes, dan catatan skor resmi, yang dikirim langsung oleh ETS kepada institusi atau agensi yang ditentukan oleh peserta tes pada kartu masuk TSE. Pemabayaran tes mengharuskan peserta tes untuk mencantumkan dua penerima laporan skor resmi. Pembayaran terpisah dibebankan untuk salinan tambahan laporan skor resmi. TOEFL/TSE Office tidak akan mengeluarkan skor TSE atau informasi lain tanpa persetujuan tertulis peserta tes.

Laporan skor resmi mencakup nama peserta tes, nomor pendaftaran, negara asal, bahasa asal, tanggal lahir, tanggal tes, dan skor TSE. Informasi yang disimpan di file TSE adalah sama dengan informasi yang tercetak pada catatan skor peserta tes dan pada laporan skor resmi. Suatu laporan skor resmi hanya akan dikirim pada institusi atau agensi yang ditunjukkan kartu masuk oleh peserta tes pada hari test atau pada formulir permintaan laporan skor yang dikirimkan pada waktu sesudahnya. Skor tersebut tidak akan dikeluarkan oleh penerima institusional tanpa izin eksplisit peserta tes.

Program TSE mengakui hak privasi peserta tes atas informasi yang disimpan dalam file data atau riset yang dipegang oleh Educational Testing Service dan merupakan tanggung jawab program TSE untuk melindungi informasi dalam filenya dari pembukaan tanpa izin. Maka, ETS tidak boleh mengefax atau memberikan hasil melalui telepon pada peserta tes atau institusi.

f. Kegunaan Skor TSE

Educational Testing Service tidak menetapkan skor lulus atau gagal pada TSE. Tiap institusi atau agensi yang menggunakan skor TSE harus menentukan skor apa yang bisa diterima, bergantung pada level kemampuan bahasa komunikasi lisan yang dianggap sesuai untuk suatu keperluan tertentu. Harus dicatat bahwa skor pada TSE revisi dan tes awal berbeda maknanya. Karena tesnya berbeda, tidak ada hubungan antar skor antara kedua alat pengukur. program TSE telah mempersiapkan TSE Standar-Setting Kit untuk membantu institusi dan agensi dalam menentukan standar skor untuk tes terrevisi.

Program TSE telah mengembangkan suatu "TSE Sampel Response Tape" sebagai suplemen untuk manual ini. Kaset audio sepanjang 30 menit ini berisi respon sampel pilihan dari TSE revisi dan dimaksudkan untuk memberi pengguna skor pemahaman yang lebih baik mengenai level-level keefektifan komunikasi yang diwakili oleh skor TSE tertentu. Kaset ini mencakup beberapa sampel tuturan yang diambil dari penutur bahasa Inggris non asli dari berbagai latar belakang bahasa asal. Sampel tuturan tersebut mewakili beragam level kemampuan bahasa Inggris lisan yang diambil dari skala rating TSE dan disusun dari skor tinggi ke skor yang rendah.

1) Panduan penggunaan skor tes TSE

Panduan berikut diberikan untuk membantu institusi dalam penafsiran dan penggunaan skor TSE.

- ◆ Gunakan skor TSE hanya sebagai suatu ukuran kemampuan untuk berkomunikasi secara lisan dalam bahasa Inggris. Jangan menggunakannya untuk memperkirakan performa akademik atau kerja.
- ◆ Dasarkan evaluasi potensi pelamar untuk kerja akademik atau performa kerja yang sukses pada semua informasi relevan yang tersedia dan sadari bahwa skor TSE hanyalah salah satu indikator kemampuan untuk bekerja secara efektif dalam konteks akademik atau profesi tertentu.
- ◆ Pertimbangkan jenis-jenis dan level bahasa lisan dalam bahasa Inggris yang diperlukan pada beragam level studi dalam disiplin akademik yang berbeda atau dalam tugas profesi yang beragam. Juga pertimbangkan sumber-sumber yang tersedia pada institusi untuk meningkatkan kemampuan bertutur bahasa Inggris dari penutur non asli.
- ◆ Pertimbangkan bahwa skor peserta tes didasarkan pada suatu rekaman 20 menit yang mewakili sampel tuturan yang spontan.
- ◆ Tinjau kembali skala rating TSE dan TSE Sampel Response Tape. Skala terkait ada di lampiran B dan kaset bisa dipesan dari TSE.
- ◆ Lakukan suatu studi validitas lokal untuk menjamin bahwa skor TSE yang dibutuhkan oleh institusi tersebut sudah tepat.

Merupakan suatu hal yang penting untuk mendasarkan evaluasi dari potensi performa kandidat internasional pada semua informasi relevan yang ada, bukan semata-mata pada skor TSE. TSE mengukur kemampuan bahasa komunikasi lisan seseorang dalam bahasa Inggris, namun tidak mengukur skill menyimak, membaca, atau menulis dalam bahasa Inggris. Tes TOEFL dan TWE bisa digunakan untuk mengukur skill-skill tersebut.

Keefektifan komunikasi lisan umum hanyalah salah satu dari banyak kualitas yang diperlukan untuk performa akademik atau pekerjaan yang berhasil. Kualitas lainnya bisa mencakup penguasaan bidang masalah, skill interpersonal, dan minat di bidang profesi terkait. TSE tidak memberikan informasi mengenai kecerdasan, motivasi, penguasaan bidang masalah atau bidang bahasan, kemampuan mengajar, atau adaptasi budaya, yang semuanya itu mungkin mempunyai peran penting pada kemampuan untuk bekerja secara efektif dalam situasi tertentu.

Sebagai bagian dari tanggung jawab umum untuk tes yang dihasilkannya, program TSE berkepentingan dalam interpretasi dan penggunaan skor TSE oleh institusi penerima catatan skor. TSE Program Office mendorong institusi per-individu untuk meminta bantuannya untuk masalah-masalah yang terkait dengan penggunaan skor TSE secara tepat.

2) Skor nonstandar

TSE Program Office merekomendasikan penggunaan metode-metode alternatif untuk mengevaluasi kemampuan bahasa Inggris pada individu-individu yang tidak bisa menjalankan tes TSE dengan kondisi standar. Kriteria seperti catatan akademik masa lalu, rekomendasi dari guru bahasa atau orang lain yang mengetahui kemampuan bahasa Inggris pelamar, dan/atau wawancara pribadi adalah hal yang disarankan sebagai pengganti skor TSE.

Seperti dinyatakan sebelumnya, TSE Program Office akan mengadakan persiapan khusus untuk mengadakan tes di bawah kondisi non standar untuk individu-individu yang cacat. Karena keadaan individual dari pelaksanaan tes non standar sangatlah beragam, TSE Program Office tidak mampu membandingkan skor-skor yang didapat dengan pelaksanaan tes khusus tersebut dengan skor yang didapat dengan tes standar. Maka TSE Program Office pun memberitahu institusi bahwa tes non standar tersebut mungkin tidak memberikan ukuran valid untuk kemampuan bahasa komunikatif lisan peserta tes, walaupun kondisinya dirancang untuk meminimalkan setiap efek buruk dari kecacatan peserta tes terhadap performa tes.

Selain itu, laporan skor resmi dan catatan skor peserta tes akan menunjukkan bahwa skor-skor tersebut didapatkan di bawah kondisi non standar. Tiap penerima skor akan juga dikirim suatu catatan penjelasan yang menekankan bahwa karena tidak terdapat data perbandingan untuk skor yang didapat di bawah kondisi pengujian nonstandar, maka skor tersebut harus digunakan secara hati-hati.

g. Riset Masa Depan dan Pengumpulan Data Statistik

Serangkaian rencana riset dan aktivitas pengumpulan data yang sedang berlangsung dan terkait dengan TSE terrevisi akan dilakukan untuk menyentuh masalah yang penting bagi program TSE, peserta tes, dan pengguna skor.

- ◆ Program TSE sedang melakukan suatu studi kesesuaian untuk menjelaskan performa peserta-peserta terpilih pada tes awal dan tes TSE revisi. Karena kedua tes berbeda dalam kandungan, format, dan rancangan skor, pendekatan kesesuaian ini akan menjelaskan korespondensi antara range (kisaran) skor pada kedua alat ukur.
- ◆ Selama tahun pertama penerapan tes TSE revisi, staf analisa statistik akan mengumpulkan informasi lebih jauh yang terkait dengan karakteristik statistik dari performa peserta tes dan tes sendiri.
- ◆ Pada tahun 1996 suatu survey pengguna skor TSE akan dilakukan untuk memastikan tingkat kepuasan dengan tes revisi dan untuk menentukan persyaratan/keperluan skor apa yang ditetapkan pada beragam tipe institusi.
- ◆ Suatu proposal riset, "Validating the Revised Test of Spoken English Against A Criterion of Communicative Success," telah dipresentasikan untuk dipertimbangkan oleh TOEFL Research Committee.

h. Speaking Proficiency English Assesment Kit (SPEAK)

Baru-baru ini program TSE menawarkan Paket Penilaian Kemampuan Bertutur Bahasa Inggris (SPEAK), yang memungkinkan para institusi untuk menerapkan bentuk TSE awal yang tidak dipakai lagi untuk tujuan evaluasi lokal. SPEAK bisa digunakan untuk pemilihan pihak yang akan dipekerjakan sebagai asisten pengajar atau dalam kapasitas yang lain. SPEAK ini juga bisa digunakan oleh program bahasa Inggris intensif untuk menempatkan siswa-siswa mereka pada level yang tepat.

SPEAK berisi semua bahan yang dibutuhkan untuk menyusun dan menerapkan suatu program pengetesan lokal. Paket ini mencakup suatu manual pelatihan rater self-instructional, yang menjelaskan cara untuk mengadakan tes dan cara untuk menggunakan skor-skor kedapat-dimengertian dan diagnostik. Juga tercakup di dalamnya adalah suatu set kaset pelatihan rater yang memberikan contoh-contoh performa peserta yang sebenarnya dalam merespon pertanyaan tes dan suatu penjelasan untuk peratingan yang diberikan untuk tiap respon. Kaset pengetesan rater memungkinkan rater untuk menskor serangkaian kaset jawaban tes lengkap dan untuk membandingkan skor-skor dengan skor sebenarnya yang diberikan oleh spesialis pengetesan bahasa ETS.

SPEAK ini juga meamsukan 30 buku tes yang bisa digunakan kembali (formulir 1) untuk peserta tes, 30 buku pegangan peserta tes dengan pertanyaan-pertanyaan sampel, satu kaset pelaksanaan tes sebenarnya, dan satu paket lembaran peratingan yang akan digunakan untuk tugas penilaian dan penghitungan skor. Selain SPEAK basis, yang mencakup Bentuk Test 1, edisi lain tes SPEAK juga tersedia. Formulir tambahannya memungkinkan untuk mengevaluasi ulang bahasa Inggris lisan dari individu-individu yang telah dites sebelumnya dengan tes SPEAK awal. Tes SPEAK awal ini hanya tersedia sampai pada akhir tahun 1995.

Untuk mengejar kualitas, SPEAK akan direvisi dan disesuaikan dengan format yang digunakan dalam TSE baru. SPEAK revisi akan diluncurkan pada musim semi tahun 1996. Sehubungan dengan hal itu, ada beberapa rencana untuk membuat tiga tipe bahan yang tersedia untuk para pengguna SPEAK:

- ◆ Satu paket pelatihan rater, yang akan mencakup kaset pelatihan rater dan alasan tertulis untuk penilaian yang diberikan pada tiap respon
- ◆ Dua bentuk tes
- ◆ Satu paket tes latihan, yang akan memungkinkan peserta tes agar terbiasa dengan format tes.

SPEAK melayani pembelian langsung oleh institut bahasa Inggris yang berafiliasi dengan universitas, organisasi pengetesan agensi atau institusional, program bahasa Inggris yang intensif, dan organisasi dan instansi lain yang melayani program pendidikan private maupun publik. Suatu hal yang penting untuk diingat adalah bahwa SPEAK dirancang hanya untuk penggunaan internal.

3. *The American College Testing (ACT)*

a. Pengenalan atau Pengertian Tentang ACT

Program Penilaian ACT (AAP) merupakan sitem pengumpulan, pengolahan, dan pelaporan data yang komprehensif dan dirancang untuk membantu siswa mengembangkan rencana postsecondary dan membantu lembaga pendidikan post secondary untuk mengembangkan program-program yang disesuaikan dengan kebutuhan dan karakteristik para pendaftarnya. Program ini membantu siswa mengidentifikasi dan mempertimbangkan banyak pilihan pendidikan post secondary dengan melalui proses bimbingan. APP membantu lembaga pendidikan post secondary dalam merencanakan

program instruksional maupun ekstrakurikuler. Selain itu, APP membantu sekolah-sekolah post secondary untuk mengevaluasi program-program yang dirancang untuk siswa yang sudah terikat dengan perguruan tinggi tertentu.

AAP terdiri dari: (1) empat test perkembangan pendidikan; (2) kuisisioner mengenai mata pelajaran dan nilai di SMA; (3) kuisisioner mengenai aspirasi karir dan pendidikan siswa, aktivitas ekstra kurikuler, dan kebutuhan pendidikan luar biasa; dan (4) suatu daftar minat. Test tersebut dilaksanakan dalam kondisi yang distandarkan, dan komponen-komponen lain dilengkapi saat siswa mendaftar untuk mengikuti tes.

Asumsi fundamental yang mendasari perkembangan tes AAP dalam perkembangan pendidikan adalah bahwa tes AAP mengukur, selangsung mungkin skill dan kemampuan yang menunjukkan kesiapan siswa untuk studi level perguruan tinggi. Dan karena performa yang berhasil di perguruan tinggi melibatkan fungsi intelektual level tinggi dalam disiplin lintas sektor yang luas, Test AAP menekankan pada kemampuan-kemampuan seperti reasoning, pemecahan masalah, dan evaluasi kritis atas bahan tertulis.

AAP dirancang pada akhir 1950-an untuk melayani kebutuhan universitas-universitas negara bagian yang besar; college junior, negara bagian, dan kota; serta sebagian besar college swasta kecil dan college keagamaan di Amerika Serikat. Walaupun pada saat itu kebanyakan lembaga-lembaga ini tidak terlibat dalam penerimaan yang selektif, mereka berusaha memberikan program-program pendidikan yang berkualitas untuk lulusan SMA yang diterima. Pertumbuhan populasi siswa yang ingin masuk perguruan tinggi pada akhir 1960-an membuat perlu bagi kebanyakan perguruan tinggi untuk mendapatkan data obyektif dan bermakna untuk mendasari keputusan tentang siswa terkait. Lembaga-lembaga dengan kebijaksanaan pemasukan terbuka memerlukan informasi yang berguna dalam membimbing siswa dalam memilih program dan penempatan jurusan. Sekolah-sekolah dengan kebijaksanaan kebijaksanaan pemasukan yang selektif memerlukan informasi yang berguna dalam mengidentifikasi siswa-siswa yang paling mungkin untuk mendapatkan keuntungan jika masuk dalam program pendidikan mereka. Maka, lembaga-lembaga tersebut membutuhkan serangkaian tes dan layanan informasi yang menyediakan suatu deskripsi yang komprehensif mengenai kebutuhan dan kemampuan pendidikan siswa mereka. Selama bertahun-

tahun, AAP telah berkembang untuk mencerminkan kebutuhan yang berubah-ubah baik dari siswa maupun dari perguruan tinggi.

Siswa merupakan titik fokus pada proses penerimaan perguruan tinggi. Karena itu, AAP dirancang untuk memberi siswa informasi yang berguna di dalam perencanaan pendidikan dan karir mereka. ACT juga memberi lembaga post secondary jangkauan luas informasi dan layanan yang bisa berguna di dalam pembuatan keputusan matang mengenai calon dan siswa yang diterima. Informasi ini digunakan untuk menyederhanakan proses penerimaan, untuk memberikan layanan saran akademik, dan untuk membuat keputusan mengenai penempatan dan pemilihan jurusan. Informasi ini juga diberikan pada SMA-SMA sehingga mereka bisa memberi saran dan membimbing siswa dalam persiapannya untuk kehidupan setelah SMA.

1) Tujuan APP

APP mempunyai banyak kegunaan (ACT, 1982a). APP membantu siswa mengidentifikasi dan mengembangkan rencana yang realistis untuk pencapaian sasaran karir dan pendidikan saat mereka pindah dari pendidikan secondary ke post secondary. Informasi AAP memberi gambaran siswa kepada lembaga post secondary sebagai orang-orang dengan pola perkembangan, pencapaian dan kebutuhan pendidikan yang unik. SMA menggunakan data AAP dalam pemberian saran dan bimbingan akademik. Perguruan tinggi menggunakan hasil AAP dalam proses penerimaan dan dalam menyesuaikan siswa dengan program instruksional dan ekstrakurikuler. Banyak agensi yang pemberi beasiswa, pinjaman, dan tipe-tipe bantuan keuangan lain pada siswa yang mengkaitkan bantuan tersebut dengan kualifikasi akademik siswa tersebut. Banyak agensi negara bagian melakukan program penghargaan untuk para penerima beasiswa negara bagian. Data AAP diberikan untuk tujuan-tujuan ini. Baru-baru ini, beragam agensi federal dan negara bagian serta beberapa konsorsium pendidikan telah menjadi bergantung pada data AAP sebagai elemen dalam evaluasi pendidikan dan aktivitas perencanaan keseluruhan mereka.

2) Populasi yang Dilayani oleh AAP

Lebih dari satu juta siswa ikatan perguruan tinggi yang berpartisipasi dalam AAP tiap tahunnya. Hampir 2.900 lembaga post secondary (termasuk agensi beasiswa, sistem pendidikan negara bagian, universitas publik dan swasta, perguruan tinggi empat-tahun

(sekolah tinggi), college junior dan community, sekolah keperawatan, dan sekolah teknik) mengharuskan atau menyarankan agar para pendaftar menyerahkan hasil AAP.

Untuk kebanyakan siswa, pendidikan post secondary mereka dimulai segera setelah mereka selesai SMA. Umumnya siswa mengambil tes AAP pada tahun ketiga atau keempat di SMA atau segera setelah mereka mengambil program matrikulasi pada suatu lembaga post secondary. Maka, kebanyakan siswa yang mengambil AAP berumur antara 16 sampai 20 tahun.

Secara historis, ACT telah memberi saran pada para siswa untuk mengambil tes AAP setelah mereka menyelesaikan sebagian besar pelajaran yang tercakup oleh tes ini. Dengan adanya kurikulum yang dimiliki oleh kebanyakan SMA dan jalur studi yang diikuti sebagian besar siswa, titik waktu ini biasanya dicapai pada musim semi tahun ketiga. Namun hal ini bervariasi dari siswa ke siswa dan dengan keempat area akademik yang diukur oleh tes AAP.

b. Deskripsi Instrumen Penilaian

Bagian pengumpulan data dari AAP terdiri dari empat instrumen: (1) empat test perkembangan pendidikan; (2) kuisioner mengenai mata pelajaran dan nilai di SMA; (3) kuisioner mengenai aspirasi karir dan pendidikan siswa, aktivitas ekstra kurikuler, dan kebutuhan pendidikan luar biasa; dan (4) suatu daftar minat. Rangkaian tes tersebut menghasilkan skor untuk empat area akademik (Bahasa Inggris, Matematika, IPS, dan IPA), dan suatu skor gabungan. Bagian Mata Pelajaran/Nilai SMA dalam AAP mengumpulkan nilai SMA yang dilaporkan sendiri oleh para siswa. Bagian Profil Siswa berisi pertanyaan-pertanyaan yang terkait dengan rencana dan kebutuhan karir dan pendidikan siswa, minat dan pencapaian bidang ekstrakurikuler, dan karakteristik latar belakangnya. Analisa pilihan siswa untuk enam area karir (sains, seni, layanan sosial, kontak bisnis, operasi bisnis, dan teknik) berasal dari respon siswa terhadap item-item dalam Daftar Minat ACT.

Siswa yang mengambil tes perkembangan pendidikan pada salah satu tanggal tes nasional melengkapi ketiga instrumen yang lain sebagai bagian dari prosedur pendaftaran. Siswa yang lain (misal yang dites sesudah tanggal tersebut) mungkin tidak bisa melengkapi ketiga instrumen tersebut.

1) Tes Perkembangan Pendidikan

Tes Penggunaan Bahasa Inggris adalah tes selama 40 menit dengan 75 item yang mengukur pemahaman siswa pada konvensi standar bahasa Inggris tertulis yang mencakup penjeadaan, tata bahasa, struktur kalimat, pilihan kata dan gaya, logika dan organisasi tulisan. Tes ini menekankan pada analisa jenis tulisan eksposisi efektif yang akan ditemui dalam banyak kurikulum post secondary, bukannya hapalan tentang aturan-aturan dan tata bahasa. Tes ini terdiri dari beberapa bacaan prosa dengan bagian-bagian tertentu yang digaris bawah dan diberi nomer. Untuk tiap bagian yang digaris bawah, respon-respon alternatif, termasuk Tidak Ada Perubahan, diberikan pada siswa. Siswa harus menentukan alternatif yang mana yang paling tepat dengan konteks dalam bacaan.

Tes Penggunaan Matematika terdiri dari 40 item selama 50 menit yang mengukur pencapaian siswa dalam matematika. Tes ini menekankan pada solusi masalah kuantitatif praktis yang dicakup dalam banyak pelajaran post secondary dan memasukan pensampelan teknik matematika yang dicakup dalam pelajaran SMA. Tes ini menekankan pada reasoning kuantitatif, bukan pada hapalan rumus, pengetahuan teknik, atau skill penghitungan. Tiap item dalam tes ini menampilkan pertanyaan dengan beberapa alternatif jawaban.

Tes Bacaan Studi Sosial adalah tes selama 35 menit dengan 52 item yang mengukur komprehensi, reasoning analitis dan evaluatif, dan skill pemecahan masalah yang diperlukan dalam studi sosial. Terdapat dua jenis item: Yang pertama di didasarkan pada wacana bacaan, kedua pada latar belakang umum atau informasi yang banyak diajarkan di pelajaran studi sosial di SMA. Semua item adalah pilihan ganda. Item-item yang didasarkan pada bacaan membutuhkan tak hanya skill pemahaman bacaan, namun juga kemampuan untuk menarik kesimpulan, untuk menyelidiki saling keterkaitan dan pentingnya ide dalam suatu bacaan, untuk memperluas ide dalam data grafik atau eksperimen, untuk melakukan deduksi dari mode, gaya, dan bias reasoning penulisnya. Item informasi tertentu meminta siswa untuk menerapkan apa yang telah mereka pelajari dalam pelajaran studi sosial SMA pada masalah-masalah serupa baru yang sering ditemui.

Tes Bacaan IPA adalah tes selama 35 menit dengan 52 item yang mengukur penafsiran, analisa, evaluasi, reasioning kritis, dan skill pemecahan masalah yang dibutuhkan dalam IPA. Terdapat dua tipe item, pertama didasarkan pada bacaan, yang

kedua didasarkan pada informasi tentang sains. Semua item adalah pilihan ganda. Bacaan-bacaan tersebut berkaitan dengan beragam topik dan masalah sains. Deskripsi mengenai eksperimen sains dan diskusi tentang teori-teori terbaru merupakan format yang paling umum. Item-item tersebut mengharuskan siswa untuk memahami dan membedakan tujuan-tujuan eksperimen tersebut, untuk menyelidiki hubungan logis antara hipotesa eksperimen dengan generalisasi yang bisa ditarik dari eksperimen tersebut, untuk memperkirakan efek-efek penerapan konsep yang disajikan dalam bacaan terkait pada situasi yang baru, untuk mengusulkan cara alternatif untuk melaksanakan eksperimen tersebut, dan untuk menilai nilai praktis dari ide dan teori yang disajikan dalam suatu bacaan. Item informasi tertentu meminta siswa untuk menerapkan apa yang telah mereka pelajari di SMA pada masalah serupa baru yang umum dijumpai. Walaupun item-item tersebut membutuhkan pemahaman fakta-fakta yang penting, item tersebut hanya memerlukan penghitungan aritmetika dan aljabar yang sedikit.

2) Bagian Informasi Pelajaran/Nilai SMA

Informasi mengenai pelajaran yang telah diterima dan nilai yang didapat dikumpulkan pada bagian ini dalam AAP Registration Folder. Para siswa diminta untuk menunjukkan apakah mereka telah mengambil atau merencanakan mengambil 30 pelajaran yang berbeda. Mereka juga diminta untuk mencatat nilai terbesar yang pernah diterima dalam setiap akhir masa belajar. Untuk meminimalkan kesalahan dalam data dilaporkan oleh siswa tersebut, para instruksi memberi saran pada para siswa untuk mengacu pada laporan atau transkrip nilai. Mereka juga diminta untuk menandatangani suatu pernyataan yang menyatakan keakuratan nilai dilaporkan.

3) Bagian Profil Siswa

Bagian Profil Siswa SPS dalam AAP Registration Folder meminta siswa untuk mensuplai informasi mengenai latar belakang, minat, kebutuhan, dan rencana mereka. Bagian ini dirancang untuk membantu siswa memikirkan masa depan pendidikan mereka dan untuk membantu lembaga post secondary untuk menyesuaikan kebutuhan siswa dengan program yang mereka tawarkan. Siswa diberitahu bahwa informasi mengenai latar belakang ras/etnik, bahasa asli, status pernikahan, agama, dan cacat fisik hanya akan dikeluarkan dengan seizin mereka pada lembaga-lembaga yang memintanya sesuai

dengan peraturan federal, dan bahwa mereka tidak diharuskan untuk memberikan informasi ini.

SPS juga meminta hampir semua informasi yang umumnya diminta perguruan tinggi dalam isian pendaftaran mereka. Siswa merespon pada pertanyaan mengenai rencana penerimaan; usulan jurusan pendidikan; level tingkat pendidikan yang dicari; kebutuhan perumahan perguruan tinggi; butuhnya bantuan dalam meningkatkan skill dalam belajar, menulis, membaca atau matematika; serta minat pada mata kuliah dan penempatan lanjutan. Sebagian lembaga menggunakan catatan AAP sebagai sistem penerimaan total (ACT, 1982a), sehingga menghilangkan pendaftaran masuk dan/atau transkrip nilai yang terpisah.

SPS juga mencakup: data demografis, informasi SMA, termasuk jumlah tahun yang digunakan untuk mempelajari pelajaran khusus di SMA; dan pencapaian prestasi yang menonjol di SMA di area kepemimpinan, musik, pidato, seni, menulis, sains, atletik, pengalaman kerja, dan layanan masyarakat.

Walaupun siswa mengisi SPS secara sukarela, kebanyakan menyadari pentingnya pemberian informasi tersebut untuk perguruan tinggi dan benar-benar mengisinya. akhir-akhir ini tercatat suatu rate respon yang lebih besar dari 95% (ACT 1987a). Pertanyaan yang membentuk SPS dan instruksi untuk menjawabnya disediakan di Lampiran B manual ini.

4) Angket Minat ACT

Angket minat ACT merupakan suatu instrumen dengan 90 item yang dirancang untuk mengukur pilihan tugas-kerja siswa dalam enam bidang yang komprehensif, yaitu: Sains, Seni, Layanan Sosial, Kontak Bisnis, Operasi Bisnis, dan Teknik. Tiap skala terdiri dari 15 aktivitas yang terkait dengan pekerjaan di mana siswa menunjukkan tingkat kesukaan mereka dalam skala tiga angka (suka, tidak peduli, tidak suka). Skala angket minat ACT dikembangkan menyerupai enam tipe pekerjaan dan minat dari Holland (1985).

c. Prosedur Pelaksanaan dan Penskoran Instrumen Penilaian

Tiap tahun AAP dilaksanakan di seluruh USA pada lima tanggal tes pada hari Sabtu. Ada juga tanggal tes yang dijadwal teratur untuk siswa yang keyakinan agamanya melarang mereka mengambil tes di hari Sabtu. Pelaksanaan khusus bisa diatur untuk

siswa yang berada dalam penjara, orang cacat, dan yang tinggal terlalu jauh dari pusat tes sehingga tidak memungkinkan melaksanakan tes sesuai jadwal. AAP juga dilaksanakan empat kali setahun di negara-negara di seluruh dunia. Personel yang sedang bertugas di markas militer Amerika bisa melaksanakan AAP melalui DANTES (Defense Activity for Nontraditional Education Support). Tanggal dan lokasi tes nasional (termasuk lokasi tidak pada hari Sabtu) dipublikasikan setiap tahun dalam tiga dokumen *Registering for the ACT Assessment* (ACT, 1987e), *The ACT Assessment Counselor's Handbook* (ACT, 1987j), dan *Using the ACT Assessment on Campus* (ACT, 1987l). ACT berusaha menyediakan cukup pusat tes untuk mengetes semua siswa yang ingin berpartisipasi dalam AAP. Tanggal tes, prosedur pendaftaran, pembayaran dan batas waktu untuk pusat-pusat tes di negara lain dipublikasikan dalam *Taking the ACT Assessment for Students Outside the 50 United States* (ACT, 1987i).

Juga terdapat program pengetesan di kampus untuk lembaga *post secondary* yang berpartisipasi dalam AAP yang mempunyai siswa yang diterima atau terdaftar namun belum mengambil tes AAP pada salah satu tanggal tes nasional. Tipe tes ini disebut tes susulan (residual testing).

Pendaftaran reguler. Untuk mendaftar AAP, siswa harus mendapatkan satu paket pendaftaran dari konselor SMA, kantor penerimaan perguruan tinggi, atau langsung dari ACT. Selanjutnya, mereka melengkapi empat halaman National Registration Folder dan mengirimkannya dan membayar biayanya pada ACT.

Paket pendaftaran siswa dikirimkan pada konselor SMA dan pada perguruan tinggi biasanya pada awal Agustus. Tiap paket berisi National Registration Folder, amplop balasan yang sudah diberi alamat, dan salinan *Registering for the ACT Assessment* (ACT, 1987e), yang memberikan instruksi pendaftaran secara rinci. *Taking the ACT Assessment for Students Outside the 50 United States* (ACT, 1987i) merupakan dokumen tersendiri yang memberikan informasi mengenai cara mendaftar AAP di negara lain.

Selain bagian-bagian untuk data identifikasi (seperti nama, alamat, tanggal lahir, dan nomor Jaminan Sosial), National Registration Folder berisi Bagian Informasi Pelajaran/Nilai SMA, Bagian Profil Siswa, dan Angket Minat ACT.

Batas waktu pendaftaran reguler kira-kira adalah empat minggu sebelum tanggal tes. Pilihan pendaftaran yang terlambat bisa didapatkan dengan tambahan biaya, dengan

batas waktu yang umumnya ditentukan pada sembilan ahri kerja sebelum tanggal tes nasional.

Jika seorang siswa mempunyai cacat yang tak akan menghalanginya untuk mengambil tes AAO pada suatu pusat tes nasional namun tetap membutuhkan persiapan khusus di pusat tes tersebut, siswa tersebut harus memberitahu ACT Test Administration Department (secara tertulis) pada saat pendaftaran. ACT kemudian membuat persiapan khusus untuk mengetes siswa tersebut jika pusat tes terkait bisa menyediakan fasilitas dan personel yang diperlukan dan jika persiapan tersebut tak akan mengganggu peserta tes yang lain. Siswa tersebut harus menggunakan booklet tes tercetak reguler dan melaksanakan keempat tes dalam waktu yang sudah disediakan. Permintaan untuk persiapan pada pusat pelaksanaan tes nasional reguler umumnya mencakup:

- 1) Pendaftar cacat pendengaran yang menggunakan penerjemah selama waktu petunjuk lisan diberikan pada para peserta tes.
- 2) Meja yang sesuai disediakan untuk peserta tes yang menggunakan kursi roda.
- 3) Peserta tes yang diabetes yang harus membawa makanan ke dalam pusat tes.

Selama satu tahun akademik terakhir, ACT menerima 121 permintaan persiapan khusus pada lima tanggal tes nasional. Permintaan tersebut umumnya dari orang cacat pendengaran yang meminta penerjemah untuk petunjuk lisan. Kesemua permintaan tersebut bisa dilayani dengan memuaskan.

Pengetesan khusus tersedia untuk pesetrt tes dengan kemampuan fisik atau perseptual yang tidak bisa melaksanakan tes AAP pada pusat tes nasional atau orang yang tidak bisa mengambil tes dalam waktu yang sudah disediakan dengan menggunakan booklet tes tercetak reguler. Pengetesan khusus juga tersedia untuk peserta tes lain yang memenuhi salah satu kriteria berikut: menginap di rumah sakit atau ada di dalam lembaga pasyarakatan pada tanggal tes terjadwal, dibatasi oleh peraturan agama untuk tidak melakukan tes di hari sabtu jika tak ada pusat tes non-sabtu yang didirikan dalam radius 50 kilometer dari rumah mereka, atau tempat tinggal di negara di mana ACT tidak mempunyai satu pusat tes.

d. Pelaksanaan Tes Perkembangan Pendidikan

Untuk bisa diizinkan masuk pusat tes, siswa harus mempunyai tiket izin masuk pusat tes ACT atau pengesahan tertulis lain dari ACT. Siswa juga harus memperlihatkan

identifikasi foto. Jika siswa tidak mempunyai SIM atau kartu siswa dengan foto terbaru, konselor SMA siswa tersebut bisa mempersiapkan surat identifikasi sebelum tanggal tes. Surat ini harus bertanda sekolah, yang ditanda tangani oleh siswa dengan kehadiran konselor, dan ditandatangani dan diberi tanggal oleh konselor. Siswa harus membawa pernyataan tersebut ke pusat tes di mana dia diharuskan menandatangani dengan kehadiran staf pusat tes. Siswa yang tidak menunjukkan identitas foto atau surat konselor yang bisa diterima tidak diizinkan melaksanakan tes.

Jadwal Pelaksanaan. Siswa diperintahkan untuk datang di pusat tes pukul 08:15 pagi pada tanggal tes di mana mereka terdaftar. Pelaksanaan tes dimulai sekitar pukul 08:15. Kira-kira 15 menit digunakan untuk instruksi, pencatatan nama pada lembar jawaban, dll. Pelaksanaan tes sebenarnya dimulai dengan Tes Penggunaan Bahasa Inggris (40 menit) diikuti tes Penggunaan Matematika (50 menit). Setelah istirahat 10 sampai 15 menit, dilaksanakan Test Bacaan IPS (35 menit) dan Tes Bacaan IPA (35 menit). Terakhir, peserta tes diminta untuk merespon suatu kuisioner singkat mengenai kondisi tes dan aktivitas persiapan tes; pada beberapa tanggal tes, peserta merespon pada lima tes bukannya empat. Salah satu dari kelima tes berisi item-item yang sedang dipraktekan untuk penggunaannya di masa datang. Pengetesan biasanya selesai pada siang hari.

ACT memberikan panduan khusus untuk mempertahankan keseragaman kondisi tes di semua pusat tes. Supervisor dan asisten pengawas yang berkualitas melaksanakan tes AAP pada pusat-pusat tes ini, yang biasanya berlokasi di SMA-SMA atau perguruan tinggi.

Untuk memberikan kondisi pengetesan terbaik untuk semua siswa, ACT telah membentuk suatu prosedur dengan tiga bagian yang mengevaluasi pelaksanaan dan keamanan tes:

- 1) Pada akhir kebanyakan sesi pengetesan, siswa diminta untuk merespon beberapa pertanyaan mengenai kondisi tes.
- 2) "Supervisor's Comment Sheet" mengundang para staf pusat tes agar memberikan saran pada ACT mengenai bahan tes dan operasi pusat tes.
- 3) Personel dari ACT secara berkala mengobservasi operasi pusat-pusat tes tertentu.

Para supervisor tes untuk semua program ujian yang disponsori ACT diberi salinan *ACT test Administration Handbook* (ACT, 1986a) dan diinstruksikan untuk membawanya

pada setiap ujian yang disponsori ACT. Dokumen ini memberikan instruksi rinci mengenai semua aspek pelaksanaan tes. Dokumen kedua, *Supervisor's Manual of Instructions* (ACT, 198h), memberikan tambahan instruksi khusus untuk AAP. Di antara prosedur standar lain, manual ini mencakup instruksi dan jadwal hari tes yang rinci untuk dibacakan pada semua peserta tes. Instruksi tersebut harus dibacakan tanpa menyimpang dari teks yang sudah ditentukan untuk menjaga kondisi tes yang standar. banyak prosedur tes tambahan ditentukan secara rinci dalam *Administration Handbook* dan *Supervisor's Manual*.

e. Prosedur Penskoran

Alat pen-scan-an elektronik digunakan untuk menskor semua bagian AAP, sehingga meminimalkan kemungkinan kesalahan penskoran. Jika seorang siswa yakin bahwa suatu kesalahan penskoran telah terjadi, dan siswa meminta secara tertulis maka ACT akan menskor langsung lembaran jawaban tadi secara gratis. Seorang siswa mungkin bisa diatur untuk hadir saat penskoran tersebut dengan mengkontak salah satu kantor regional ACT, tetapi harus membayar semua biaya tambahan yang mungkin muncul dalam menyelenggarakan layanan khusus ini, Kerahasiaan yang ketat untuk catatan tiap siswa tetap dipertahankan.

Untuk tanggal-tanggal tes tertentu (yang ditentukan dalam *Registering for the ACT Assesment*, ACT, 1987e), tiap peserta tes bisa mendapatkan (dengan sedikit biaya) salinan pertanyaan tes yang digunakan untuk menentukan skornya. Salinan jawaban pertanyaan siswa, daftar jawaban yang benar untuk pertanyaan tersebut, nilai mentah yang digunakan untuk menghitung skor dilaporkan, dan satu tabel untuk mengubah nilai mentah menjadi skor standar yang dilaporkan. Bahan-bahan ini tersedia hanya untuk siswa yang dites selama pelaksanaan reguler tes AAP pada tanggal-tanggal tes nasional tertentu.

ACT mempunyai hak untuk membatalkan skor tes siswa jika, untuk alasan tertentu, terdapat keraguan kuat mengenai validitas skor. Contohnya, jika ACT menyatakan bahwa seorang siswa telah mencontek, ACT bisa membatalkan skor tesnya. Jika ACT mempunyai keraguan kuat mengenai validitas satu set skor tes, atau alasan untuk mencurigai bahwa keanehan telah terjadi, ACT secara rutin melaksanakan penyelidikan. Dalam kasus-kasus di mana skor siswa dipertanyakan, maka akan dilakukan kajian yang

sangat cermat. Untuk semua kasus seperti itu, siswa diberitahukan mengenai pilihan-pilihan yang ada bagi mereka, termasuk prosedur banding.

Untuk keempat test content area dalam AAP (Penggunaan Bahasa Inggris, Penggunaan Matematika, Bacaan IPS, dan Bacaan IPA), skor mentah (jumlah jawaban yang benar) diubah menjadi skor standar berdasar suatu skala yang berkisar dari 1 sampai 36). Skor minimum standar untuk setiap tes adalah 1. Skor standar maksimum sedikit berbeda untuk tiap tes: Penggunaan Bahasa Inggris, 33; Penggunaan matematika, 36; Bacaan IPS, 34; dan Bacaan IPA, 35.

Skor *Composite* merupakan rata-rata keempat skor standar yang dibulatkan pada nilai bulat terdekat (0,5 dibulatkan ke atas). Skor composite minimum adalah 1, maksimum 35.

Bagian Profil Siswa. Respon pada Bagian Profil Siswa (SPS) ditabulasikan dan data hasilnya dimasukkan ke dalam laporan AAP yang disediakan untuk SMA, seperti dijelaskan dalam bagian selanjutnya manual ini.

Angket Minat ACT. Skor pada instrumen ini diberikan untuk individu yang menjawab setidaknya 10 dari 15 item pada keenam skala. Skor mentah diubah menjadi skor standar (rata-rata 50, standar deviasi 10) dan dilaporkan dalam bentuk tersebut. Ranking persentase, yang didasarkan pada sampel representatif skala nasional untuk siswa tahun keempat SMA, dilaporkan dalam bidang standard error.

f. Prosedur Pelaporan Skor

Informasi dari AAP di-agregat-kan dalam beragam cara berdasar pada penerima yang dituju dan kemudian dilaporkan. Laporan unik diberikan untuk siswa (Laporan Siswa), SMA (Laporan SMA), dan juga pada enam universitas dan agensi lain (Laporan Perguruan Tinggi) yang dipilih oleh siswa pada formulir pendaftaran. Label SMA dan Laporan Daftar SMA juga dikirimkan setelah setiap tanggal tes nasional ke SMA tempat belajar siswa terkait. Laporan-laporan ini secara singkat dijelaskan di bawah ini dan lebih rinci dalam *ACT Assesment Counselor's Handbook* (ACT, 1987j) dan *Using the ACT Assesment on Campus* (ACT, 1987l). Laporan siswa disertai dengan *Using Your ACT Assesment Results* (ACT, 1987m), satu booklet yang dirancang untuk membantu siswa dalam perencanaan pekerjaan dan pendidikannya.

1) Laporan Siswa

Laporan Siswa dirancang untuk membantu tiap siswa untuk memahami hasilnya. Jika seorang siswa menerima laporannya, laporan tersebut disertai dengan *Using Your ACT Assessment Results* (ACT, 1987m). Satu tujuan dari laporan ini adalah untuk memberikan penjelasan skor yang memadai sehingga siswa tak akan memerlukan penjelasan yang ekstensif dari konselor mereka. Maka waktu konselor dapat digunakan untuk aktivitas bimbingan yang lebih rumit, misalnya pemaduan hasil AAP dengan perencanaan pekerjaan dan pendidikan keseluruhan siswa.

Format laporan naratif yang sekarang digunakan diperkenalkan pada tahun akademik 1982-1983 dan dirancang untuk memberikan jawaban pada pertanyaan yang umumnya ditanyakan oleh siswa. Pertanyaan-pertanyaan tersebut diidentifikasi dari interpretasi hasil AAP yang direkam oleh konselor SMA yang sedang berpraktik. Selama hampir dua tahun perkembangan, pemrograman, dan pengetesan lapangan sebelum pengenalan laporan, konselor, siswa dan orang tua membantu anggota staf ACT dalam menyempurnakan narasi tersebut.

2) Laporan SMA

Jika siswa mencatat kode SMA pada National Registration folder, SMA tersebut menerima Laporan Sekolah untuk siswa tersebut. Laporan ini memberi beberapa informasi kepada konselor, di antaranya:

- a) Nilai SMA terbaru yang dilaporkan sendiri oleh siswa yang diterima sebelum tahun keempat SMA; skor standar untuk keempat tes AAP dan **Composite**-nya; dan ranking persentase skor ini berdasar pada norma-norma SMA lokal, perguruan tinggi negara bagian, dan perguruan tinggi nasional yang tersedia bagi siswa yang mengikuti AAP.
- b) Informasi yang dilaporkan sendiri oleh siswa mengenai aktivitas ekstrakurikuler SMA yang dipilih, rencana ekstrakurikuler perguruan tinggi, dan pencapaian di luar kelas.
- c) Perkiraan tanggal yang direncanakan siswa masuk perguruan tinggi, dan rencana perumahan untuk tahun pertama, Pesan khusus disediakan di sini.
- d) Respon siswa "ya" atau "tidak" untuk item 190 dalam Bagian Profil Siswa, yang memberikan izin pengeluaran data siswa ke perguruan tinggi dan agensi beasiswa melalui ACT Educational Opportunity Service.
- e) Rencana Pekerjaan dan Pendidikan saat ini, terpilih dari suatu daftar dengan 190 jurusan dan pekerjaan yang berbeda. Siswa menunjukkan tingkat kepastian rencana-

rencana ini, dan memperkirakan IPK perguruan tinggi pada tahun pertama dalam skala 4 angka.

- f) Informasi yang dilaporkan sendiri mengenai kebutuhan siswa akan bantuan khusus, minat mata pelajaran yang diujikan (credit by examination), penempatan tingkat lanjut, mata kuliah yang diinginkan (honors courses) pada tahun pertama, dan program studi independen.

g. Riset dan Layanan Informasi Terkait

Bagian ini berisi tinjauan singkat dari berbagai layanan yang diberikan oleh ACT selain laporan-laporan yang dijelaskan dalam bagian yang lalu. Lebih banyak informasi mengenai layanan- layanan ini bisa ditemukan dalam booklet *Research and Information Services* (ACT, 1987g) dan dalam manual-manual tertentu untuk tiap layanan yang diacu dalam bagian ini.

1) Layanan Profil SMA

Layanan ini diberikan pada SMA-SMA asal 40 atau lebih siswa dalam kelas yang baru lulus dan berpartisipasi dalam AAP dan melaporkan kode sekolahnya. Hanya pengetestan yang paling akhir yang dimasukkan untuk siswa yang dites lebih dari sekali. Profil ini memberikan informasi kelompok yang deskriptif mengenai siswa-siswa ini untuk digunakan dalam konseling dan dalam perencanaan dan evaluasi program. Layanan ini mencakup Laporan Profil SMA ACT, yang menyajikan tabel statistik yang menjelaskan beragam karakteristik siswa yang dites. Karakteristik berikut adalah yang diterangkan: perkembangan pendidikan, hubungan skor tes dengan persiapan akademik, sasaran dan aspirasi, pencapaian di SMA dan karakteristik tubuh siswa, evaluasi dari siswa tentang SMA lokal, dan pilihan perguruan tinggi. Laporan ini juga berisi ringkasan penemuan dan kecenderungan utama.

Pada bulan September, salinan Laporan Profil SMA ACT dikirimkan secara otomatis, tanpa bayaran, pada setiap sekolah yang memenuhi syarat. Salinan dari tiap laporan profil lokal juga dikirim pada penilik sekolah tersebut. Tercakup juga dalam pengiriman ini yaitu:

- a) Satu set chart yang harus diisi dan digunakan untuk membantu menyajikan data terlapor pada orang lain.

- b) Norma negara bagian untuk negara bagian SMA tersebut (hanya tersedia untuk negara bagian yang setidaknya 1000 siswanya dites selama tiga tahun terakhir), norma regional, dan norm nasional
- c) Salinan *College Student Profiles: Norms for the ACT Assesment* (ACT, 1987a; direvisi secara berkala dan dimasukkan hanya jika sudah direvisi tahun tersebut)
- d) Salinan *Your College-Bound Students: Interpretive Guide to the ACT High School Profile Service* (ACT, 1986d) untuk digunakan dalam menginterpretasikan tabel-tabel dalam laporan tersebut (direvisi secara berkala dan dimasukkan hanya jika sudah direvisi tahun tersebut)

2) Layanan Kesempatan Pendidikan

Layanan Kesempatan Pendidikan (EOS) ACT (ACT, 1987b) bisa membantu siswa untuk menyadari kesempatan bantuan keuangan dan pendidikan tertentu yang tampak sesuai dengan minat pendidikan, sasaran, dan kemampuan mereka, seperti tercermin dalam catatan ACT mereka. Data yang relevan dalam catatan ini mencakup nilai SMA dan pelajaran yang diambil dan dilaporkan sendiri oleh siswa, bidang jurusan studi yang diharapkan, rencana pekerjaan, sasaran gelar sarjana, kebutuhan keuangan, pencapaian prestasi di luar kelas, minat ekstrakurikuler, dan kebutuhan dan sasaran pendidikan khusus.

Lembaga-lembaga *post secondary* yang tertarik untuk menggunakan EOS mempunyai dua pilihan layanan. Satu adalah dengan menerima label pengiriman pos atau catatan komputer yang berisi informasi tentang siswa-siswa terpilih. Perguruan tinggi terkait kemudian mengirimkan informasi pada siswa mengenai program bantuan keuangan atau pendidikan yang sedang ditawarkan. Dalam pilihan keduanya, perguruan tinggi tersebut memberi ACT suatu deskripsi singkat mengenai kesempatan tertentu, dan ACT mengirim-kan deskripsi ini pada siswa terpilih yang mungkin kemudian mengontak perguruan tinggi terkait untuk mencari tambahan informasi.

Untuk kedua pilihan, ACT menelusuri file-filenya untuk mengidentifikasi siswa yang karakteristiknya memenuhi kriteria yang ditetapkan oleh perguruan tinggi atau agensi terkait. Siswa-siswa yang terpilih adalah siswa yang memberi hak pada ACT untuk mengeluarkan namanya untuk tujuan layanan ini dengan merespon "ya" pada item 190 dalam Bagian Profil Siswa. Yang berhak menggunakan EOS hanyalah Perguruan

tinggi atau agensi yang mengharuskan atau merekomendasikan nilai AAP dan mempublikasikan fakta tersebut dalam informasi mereka untuk calon mahasiswanya. Lembaga-lembaga post secondary yang menggunakan layanan ini diharuskan untuk berpartisipasi sejak awal tahun akademik karena siswa umumnya siswa seperti itu secara aktif terlibat dalam pembuatan keputusan perguruan tinggi pada saat itu. Perguruan-perguruan tinggi atau agensi membayar biaya untuk berpartisipasi dalam EOS.

3) Layanan Informasi Penerimaan

Dengan perkiraan turunnya jumlah lulusan SMA pada dekade depan, lembaga-lembaga post secondary menjadi lebih berkepentingan dengan terpeliharanya penerimaan mahasiswa dan mencari cara yang lebih efektif untuk memonitor dan mengevaluasi proses penerimaan/perekrutan mahasiswa. Layanan informasi penerimaan (EIS) dari ACT (ACT, 1987k) memberi para lembaga tersebut suatu pasar informasi yang berguna di dalam perekrutan dan keberadaan mahasiswa di perguruan tinggi tersebut.

EIS mempunyai dua komponen: Layanan Analisa Pasar dan Layanan Analisa Hasil. Data base yang dihasilkan oleh sekitar satu juta siswa yang ingin melanjutkan ke perguruan tinggi dan berpartisipasi dalam AAP setiap tahunnya digunakan dalam kedua komponen tersebut. EIS memadatkan informasi tersebut ke dalam tabel-tabel yang berfokus pada karakteristik siswa (Skor Composite ACT, jenis kelamin, penghasilan keluarga, latar belakang etnik/ras) dan faktor-faktor pilihan perguruan tinggi (ukuran dan tipe lembaga post secondary yang diinginkan, jurusan pendidikan yang direncanakan, rencana untuk tinggal di negara bagiannya atau keluar negara bagian) sesuai segment pasar.

Laporan Analisa Pasar dan Laporan Analisa Hasil menyajikan data yang diolah untuk menunjukkan karakteristik siswa yang diketahui dan faktor pilihan perguruan tinggi sesuai dengan interval Skor Composite AAP, dan sesuai jurusan pendidikan yang diinginkan.

Laporan Analisa Pasar memberikan pola rinci segmen pasar dari suatu institusi dan bisa membantu dalam mengetahui segmen pasar baru dengan potensi tinggi. Laporan Analisa Hasil memungkinkan suatu lembaga untuk mengevaluasi keberhasilannya dalam

menarik beragam tipe siswa dari segment yang teridentifikasi, dan membandingkan proporsi siswa dari suatu segmen tertentu yang mengungkapkan suatu minat pada suatu lembaga dengan proporsi yang sebenarnya mendaftar pada lembaga tersebut.

4) Layanan Riset ACT

ACT menawarkan layanan riset yang memberi pada perguruan tinggi dan universitas dengan ringkasan statistik dari data yang terkumpul dari siswa-siswa mereka. Layanan riset ACT merupakan bagian integral dari riset kelembagaan dan program belajar-sendiri pada banyak lembaga.

Layanan riset ini dirancang untuk membantu perguruan tinggi dalam riset kelembagaan mereka dan program belajar-sendiri dengan memberikan analisis data seragam melalui beragam program deskriptif, prediktif dan evaluatif. Partisipasi dalam layanan ini memungkinkan perguruan tinggi untuk mendapatkan analisis rutin dari data siswa untuk perencanaan kelembagaan dan evaluasi. Karena adanya pemrosesan massa, analisa tersebut ekonomis dan data normatif yang ekstensif bisa dikembangkan untuk tujuan komparatif dengan sedikit kerja pada bagian petugas kelembagaan.

Semua lembaga post secondary yang menerima data AAP memenuhi syarat untuk suatu layanan riset ACT, karena lembaga tersebut telah berpartisipasi dalam program pengumpulan data penting yang mensuplai data untuk analisa terkait dan dapat memenuhi persyaratan ukuran sampel dan jadwal waktu yang sudah pasti. Lembaga post secondary yang memenuhi syarat bisa berpartisipasi tanpa dipungut biaya dalam Layanan Profil Kelas dan Layanan Riset Standar atau Dasar, yang dijelaskan di bawah. Publikasi *Your College Freshmen: Interpretive Guide to ACT Research Services for Higher Education* (ACT, 1981b), memberikan bimbingan pada lembaga-lembaga yang menggunakan layanan ini.

Layanan Profil Kelas (ACT, 1986b) memberikan laporan yang komprehensif mengenai kelas mahasiswa tahun pertama yang masuk dan laporan lain dari siswa-siswa yang mengirimkan skor mereka ke perguruan tinggi terkait namun tidak mendaftar di sana. Jika perguruan tinggi tersebut menginginkannya, laporan bisa didapatkan untuk dua kelompok mahasiswa tahun pertama tertentu (diklasifikasikan berdasar perguruan tinggi) di kampus, sebagai pengganti kedua laporan profil tersebut.

Dengan menggunakan skor tes, nilai SMA, dan informasi pendidikan lain yang dikumpulkan secara rutin sebagai bagian dari AAP, laporan tersebut berisi serangkaian tabel yang menggambarkan karakteristik kunci dari kelompok siswa yang masuk, termasuk:

- kemampuan, sasaran dan aspirasi siswa
- kebutuhan tertentu yang relevan dengan layanan personel siswa
- pencapaian prestasi siswa di luar kelas
- laporan pertimbangan pilihan perguruan tinggi
- laporan karakteristik demografis
- tabulasi silang antara karakteristik siswa yang berbeda
- kecenderungan dari tahun ke tahun

Suatu ringkasan naratif dari penemuan yang penting dimasukkan ke dalam laporan tersebut.

Layanan Riset Dasar memberikan analisa hubungan antara data AAP (skor standar AAP dan rata-rata SMA) dengan nilai keseluruhan perguruan tinggi tahun pertama, dan juga persamaan prediksi untuk memperkirakan performa calon mahasiswa. Jumlah pengumpulan data yang dibutuhkan dari perguruan tinggi yang berpartisipasi hanya sedikit dan hasilnya disajikan dalam suatu laporan singkat. Partisipan yang paling sering ikut dalam layanan ini adalah perguruan tinggi kecil atau yang baru berpartisipasi, atau perguruan tinggi dengan staf yang terbatas. Persamaan prediksi yang dikembangkan dengan layanan ini digunakan untuk nilai yang diperkirakan dari siswa individual yang merupakan bagian Laporan Perguruan Tinggi, yang dijelaskan sebelumnya dalam bab ini.

Layanan Riset Standar (ACT, 1987d) memberikan hal-hal berikut pada para partisipannya: satu deskripsi kemampuan dan prestasi akademik dari kelas tahun pertama saat, dan persamaan prediksi yang digunakan dalam meramalkan performa siswa di masa datang; suatu analisa komprehensif dari informasi pra-perguruan tinggi, yang mencakup skor tes AAP, nilai SMA terlapor oleh siswa, dan jika diinginkan, informasi yang dikumpulkan dalam lingkungan lokal; data prestasi semester atau tahun pertama untuk sebanyak 9 kelompok berbeda siswa yang diidentifikasi oleh perguruan tinggi; laporan ringkasan juga dimasukkan. Nilai perguruan tinggi keseluruhan dan nilai dalam empat mata kuliah atau area subyek tertentu bisa dianalisis.

Layanan Riset Standar memungkinkan analisa yang melibatkan sampai 40 nilai mata kuliah tertentu dan 10 rata-rata nilai keseluruhan. Persamaan prediksi yang dikembangkan dalam analisa tersebut digunakan untuk melengkapi prediksi akademik pada laporan skor (Laporan Perguruan Tinggi, yang dijelaskan di atas) untuk siswa yang dites AAP pada tahun selanjutnya.

Suatu perguruan tinggi bisa juga memilih melaporkan sampai 5 alat ukur prediktif yang dikumpulkan secara lokal untuk tiap kelompok siswa. Maka, analisa layanan riset menyediakan fasilitas untuk membandingkan alat-alat ukur ini dengan alat prediksi AAP standar.

5) Divisi Layanan Pendidikan ACT

Divisi Layanan Pendidikan ACT mengelola kantor lapangan regional dan seorang profesional untuk membantu SMA, lembaga post secondary, dan agensi pendidikan untuk melakukan penggunaan data dan layanan ACT secara optimal.

Setiap konsultan regional telah mempunyai pengalaman banyak dalam komunitas pendidikan. Dia mengetahui potensi penuh layanan ACT dan bisa memberi saran mengenai cara terbaik untuk melakukan penggunaan lokal yang paling tepat dari data yang tersedia. Bantuan konsultan-konsultan ini adalah bagian dari layanan reguler ACT dan tidak meng-haruskan adanya biaya atau kewajiban dari pihak institusi.

Kantor-kantor regional ACT bisa dikontak untuk dimintai bantuan untuk hal-hal berikut:

- menggunakan data ACT untuk menyederhanakan proses penerimaan
- menggunakan data ACT untuk merekrut, menerima, dan melayani calon siswa
- menggunakan data ACT dalam bimbingan dan konseling, penempatan kelas, dan pemberian saran
- menggunakan layanan ACT untuk meningkatkan hubungan SMA dan retention siswa
- melaksanakan seminar pengembangan profesional
- menggunakan data ACT dalam operasi bantuan keuangan
- merencanakan dan menyajikan penemuan riset
- membangun sistem informasi siswa

- merencanakan dan melaksanakan workshop fakultas perguruan tinggi dan SMA

h. Prosedur Pengembangan Instrumen

Sub Bab ini akan menjelaskan prosedur yang digunakan dalam pengembangan dan revisi beragam instrumen dalam AAP. Keempat bagian sub bab ini menjelaskan prosedur yang terkait dengan empat instrumen, yaitu : (1) test perkembangan pendidikan; (2) kuisisioner mengenai mata pelajaran dan nilai di SMA; (3) kuisisioner mengenai aspirasi karir dan pendidikan siswa, aktivitas ekstra kurikuler, dan kebutuhan pendidikan luar biasa; dan (4) suatu daftar minat.

1) Tes Perkembangan Pendidikan ACT

Siklus perkembangan tes yang dibutuhkan untuk menghasilkan tiap bentuk baru Tes Perkembangan Pendidikan adalah 2,5 tahun dan melibatkan beberapa tahap, yang dimulai dengan tinjauan kembali spesifikasi tes. Tiap tahun ACT menghasilkan sejumlah bentuk tes untuk memenuhi kebutuhan tes 2,5 tahun ke depan.

a) Mempersiapkan Spesifikasi Tes

Dua jenis spesifikasi tes digunakan dalam pengembangan tes AAP: spesifikasi kandungan dan spesifikasi statistik.

Spesifikasi kandungan untuk tes AAP dikembangkan dengan bantuan konsultan pendidikan yang dikenal secara nasional. Konsultan tersebut bertemu dengan staf ACT untuk menentukan topik khusus yang akan dicakup dalam tes dan juga proporsi item untuk mengukur area topik tersebut. Cakupan kandungan dan jumlah item yang tercakup dalam tiap tes dipilih untuk mencerminkan penekanan pada kurikulum SMA dan post secondary.

Kehati-hatian tetap dilakukan untuk menjamin bahwa struktur dasar tes AAP tetap sama dari tahun ke tahun sehingga skornya bisa dibandingkan/setara, alokasi khusus item tes untuk tes tersebut juga "diselaraskan" dari secara tahunan. Tiap tahun panel konsultan dilakukan untuk meninjau kembali bentuk baru tes untuk memverifikasi keakuratan kandungan mereka dan kesesuaian kandungan tes dengan penekanan kandungan kurikulum SMA dan persyaratan mata kuliah post secondary. Lebih jauh, performa siswa di tiap area kandungan dimonitor sehingga pergeseran dalam kurikulum SMA bisa

dicatat. Informasi ini digunakan saat item dipilih untuk dimasukkan ke dalam tes. Secara berkala, survey kurikulum nasional juga dilakukan untuk mengidentifikasi pergeseran dan perubahan penekanan dalam kurikulum SMA dan post secondary.

Spesifikasi Statistik untuk tes menunjukkan level kesulitan (proporsi kebenaran) dan level diskriminasi item tes minimum yang bisa diterima (korelasi serial) dari item test yang akan digunakan.

Distribusi item kesulitan diseleksi agar tes itu akan secara efektif membedakan antara murid yang mengetahui materi dan yang tidak mengetahui materi. Tes disusun agar mempunyai kesulitan item rata-rata sekitar 0,50 dan rentang kesulitan dari 0,30 hingga 0,80.

Dengan melihat indeks diskriminasi, standar yang digunakan saat menyeleksi item untuk tes AAP adalah masing-masing item harus mempunyai korelasi biserial 0,30 atau lebih dengan nilai pada tes yang mengukur kandungan yang setara. Sebagai contoh, item Matematika harus berkorelasi dengan performa pada Tes Penggunaan Matematika.

Tes-tes yang memenuhi spesifikasi statistik bisa diharapkan dapat memberikan ketepatan pengukuran yang tinggi untuk kebanyakan siswa, sambil tetap memberikan pengukuran yang baik untuk siswa berprestasi baik yang bersaing mendapatkan beasiswa.

b) Penseleksian Penulis Item

Tiap tahun, ACT mengkontrak penulis item untuk membuat sekitar 3.300 item untuk AAP. Para penulis item adalah spesialis kandungan dalam disiplin yang diukur oleh tes AP. Kebanyakan dari mereka secara aktif terlibat dalam pengajaran. Mereka mengajar pada sejumlah level yang berbeda, dari SMA sampai universitas, dan pada berbagai lembaga, dari lembaga swasta kecil sampai lembaga publik besar. Mereka mewakili keragaman populasi dari Amerika Serikat dengan melihat latar belakang etnik, gender, dan lokasi geografis. ACT melakukan banyak usaha untuk menjamin agar para penulis untuk AAP mewakili lintas sektor pendidik di Amerika Serikat.

Sebelum diminta untuk menuliskan item-item untuk tes AAP, calon penulis item diharuskan untuk menyerahkan suatu set sampel bahan untuk tinjauan. Panduan untuk penulis item khusus untuk area kandungan terkait diberikan pada setiap penulis item. Panduan ini mencakup contoh-contoh item, dan juga memberi spesifikasi tes pada penulis item dan persyaratan ACT dengan melihat kandungan dan gaya. Selain itu dimasukkan

juga spesifikasi untuk gambaran yang adil untuk subkelompok individual, penghindaran subyek masalah yang asing untuk anggota suatu subkelompok, dan penggunaan bahasa yang melecehkan jenis kelamin tertentu.

Setiap unit yang diserahkan oleh calon penulis item dievaluasi oleh staf pengembangan tes ACT. Keputusan mengenai pengontrakan penulis item dibuat berdasar pada evaluasi tersebut.

Setiap penulis item yang dikontrak diberi tugas menghasilkan sejumlah kecil (biasanya 18 sampai 30) item pilihan ganda. Ukuran penugasan menjamin bahwa beragam bahan akan dihasilkan dan bahwa keamanan program pengetesan akan dipertahankan, karena setiap penulis item hanya akan mengetahui sebagian kecil item yang dihasilkan. Para penulis item bekerja sama dengan spesialis tes ACT yang membantu mereka dalam menghasilkan item-item dengan kualitas tinggi yang memenuhi spesifikasi tes.

Penulis item berasal dari berbagai jenis kelamin, latar belakang etnik/ras, level pendidikan dan mengajar di lembaga pengajaran yang beragam dari sudut level, besar, dan tipe lembaga terkait.

c) Konstruksi Butir

Penulis item harus menciptakan item yang penting untuk pendidikan selain logis dari sudut psikometrik. Sejumlah besar item harus di-konstruksi karena, walaupun dengan adanya penulis yang baik, banyak item tidak bisa memenuhi standar yang ditetapkan untuk item-item tes yang bisa digunakan.

Setiap penulis item menyerahkan satu set item dari suatu bidang atau area tertentu. Semua item pada Tes Penggunaan Matematika dan sebagian item pada Bacaan IPS dan Bacaan IPA berdiri sendiri (tidak terkait dengan wacana bagian). Item sisanya pada kedua bacaan dan semua item pada Tes Penggunaan Bahasa Inggris terkait dengan wacana bacaan.

Saat satu unit tes diserahkan oleh seorang penulis tes, unit tersebut ditinjau oleh staf ACT dengan melihat akurasi kandungan dan kualitas komposisi secara umum. Suatu keputusan dibuat untuk: (a) menerima unit tersebut; (b) meminta penggantian atau revisi untuk bagian-bagian unit tersebut; (c) menolak unit tersebut. Jika revisi unit tersebut diperlukan, unit tersebut ditinjau kembali saat unit tersebut diserahkan kembali.

Setelah satu unit diterima, unit tersebut diedit untuk memenuhi spesifikasi ACT dalam akurasi kandungan, panjang kata, klasifikasi item, format item, dan bahasanya. Selama proses pengeditan, semua bahan tes ditinjau untuk melihat gambaran representativitas untuk sub-sub kelompok dan penggunaan bahasa yang tepat. Unit ini ditinjau beberapa kali oleh staf ACT untuk menjamin bahwa unit ini memenuhi semua standar ACT.

Satu salinan tiap unit kemudian dikirimkan kepada konsultan yang mengkhususkan diri pada bidang tertentu yang relevan. Mereka memverifikasi keakuratan dan mengecek beragam pertimbangan mengenai kontrol kualitas. Konsultan-konsultan ini sebelumnya telah diberi pelatihan oleh staf ACT. Setiap komentar mengenai unit yang dikembalikan pada staf ACT oleh para konsultan ditinjau kembali oleh staf ACT tersebut, dan perubahan secukupnya diadakan. Jika terdapat ketidaksepakatan dalam saran para konsultan mengenai perubahan dalam kandungan, unit tersebut dikirim ke konsultan lain untuk ditinjau kembali. Jika terdapat perubahan dalam unit tersebut sebagai hasil dari komentar para konsultan, bagian-bagian yang diubah dikirimkan kembali kepada konsultan untuk memastikan agar perubahan tersebut dibuat dengan benar.

d) Uji-coba Butir

Item yang dinilai bisa diterima dalam proses peninjauan kemudian dirakit ke dalam unit-unit tryout. Satu tanggal tes nasional dipilih dan sampel siswa yang mendaftar AAP pada tanggal itu dipilih. Sampel-sampel ini, yang masing-masing terdiri dari lebih dari 200 siswa, secara hati-hati dipilih untuk mewakili populasi total peserta tes. Setiap sampel diberikan satu unit tryout dari salah satu area subyek masalah yang dicakup oleh tes AAP. Setiap peserta tes mengerjakan unit tryout tersebut selain keempat tes reguler.

e) Analisis Item dari Ujicoba

Analisa item dilaksanakan pada unit ujicoba (tryout). Untuk satu unit tertentu, sampel dibagi ke dalam kelompok rendah, menengah, dan tinggi sesuai skor individu pada tes AAP di area kandungan yang sama (yang dikerjakan pada saat yang sama seperti unit tryout). Skor "cutting" untuk ketiga kelompok adalah titik persentase 27% dan 73% dalam distribusi skor-skor tersebut.

Proporsi siswa dalam tiap kelompok yang menjawab dengan benar setiap item tryout ditabulasikan, demikian juga proporsi dalam tiap kelompok yang memilih jawaban

yang salah. Koefisien biserial dan point-biserial antar setiap skor item (benar/salah) dan total skor pada tes sejenis dalam bentuk tes (nasional) reguler juga dihitung.

Analisa item berfungsi untuk mengidentifikasi pertanyaan item tes yang efektif secara statistik. Item-item, yang terlalu sulit ataupun terlalu mudah dan yang gagal untuk membedakan siswa denn perkembangan pendidikan tinggi dan rendah seperti diukur oleh skor tes AAP reguler mereka, dihilangkan atau direvisi. Koefisien korelasi biserial dan point-biserial, dan juga perbedaan antara proporsi siswa yang menjawab benar item dalam ketiga kelompok, digunakan sebagai index kekuatan pembeda dari item pra-tes.

f) Perakitan Bentuk yang Baru

Item-item yang dinilai bisa diterima dalam proses peninjauan ditempatkan dalam suatu "pool" item. Bentuk awal dari tes AAP disusun dengan memilih dari pool ini item-item yang memenuhi spesifikasi kandungan dan statistik untuk tes-tes tersebut.

Untuk tiap tes dalam rangkaian tes, item-item untuk bentuk-bentuk baru dipilih untuk memenuhi distribusi kandungan untuk tes tersebut. Pada Tabel 2.6, frekuensi yang paling rendah untuk kedua tahun obeservasi dalam kesalahan menjawab ada pada rentang kesulitan 0,00 sampai 0,09 (jumlahnya 0). Sedangkan frekuensi paling tinggi untuk tahun 1984-85 ada pada rentang kesalahan 0,70 sampai 0,79 sebanyak 99 kali, dan untuk tahun 1985-86 ada pada rentang kesalahan 0,60 sampai 0,69 sebesar 103 kali. Selain distribusi kesulitan item, indeks pembeda (diskriminasi) item dalam bentuk korelasi biserial rata-rata yang terobservasi bisa diperkirakan untuk menunjukkan fluktuasi tertentu dikarenakan variasi dalam siswa yang diambil sebagai sampel.

Versi awal bentuk tes dihadapkan pada beberapa tinjauan untuk menjamin bahwa item-item tersebut adalah akurat dan bahwa bentuk tes keseluruhan sesuai dengan praktik penyusunan tes yang baik. Tinjauan pertama dilakukan oleh staf ACT. Item dicek dalam akurasi kandungan dan kesesuaiannya dengan gaya ACT. Item tersebut juga ditinjau untuk menjamin bahwa item bebas dari petunjuk yang bisa memungkinkan siswa yang ber-pengalaman dalam tes untuk menjawab item dengan benar walaupun mereka kurang menguasai pengetahuan dalam area subyek.

Untuk setiap item yang dipilih untuk tes Bahasa Inggris, IPS, dan IPA, terdapat empat alternatif respon yang dibuat; alternatif yang dianggap paling buruk berdasar hasil

prates dan/atau tinjauan subyek masalah akan dihilangkan. Dalam Tes Matematika terdapat lima alternatif jawaban.

Bentuk-bentuk tes nasional kemudian dikirimkan ke konsultan nasional untuk ditinjau. Untuk tes Matematika dan Penggunaan Bahasa Inggris, setiap bentuk ditinjau oleh konsultan kandungan, konsultan pengukuran, dan konsultan minoritas. Untuk tes IPA dan IPS, setiap bentuk ditinjau secara keseluruhan oleh konsultan pengukuran dan konsultan minoritas. Selain itu, konsultan kandungan meninjau bagian bentuk tes yang relevan dengan bidang keahlian mereka (misal, sejarah, ilmu politik, ekonomi, sosiologi, psikologi, biologi, kimia, fisika, astronomi, atau geologi). Konsultan-konsultan ini bukanlah individu yang sama dengan yang digunakan untuk tinjauan kandungan dari unit tryout.

Semua konsultan diberikan checklist rinci yang digunakan untuk menjamin bahwa semua bahan tes memenuhi persyaratan kandungan dan gaya dari ACT.

Dalam bagian akhir proses tinjauan, panel konsultan disusun untuk keempat tes AAP. Panel untuk tiap tes terdiri dari ahli subyek masalah dari tiap disiplin yang relevan dan juga staf ACT. Karena anggota panel membahas setiap item dengan mengacu pada semua disiplin yang sesuai dan juga pada spesifikasi tes, prosedur ini menjamin keakuratan kandungan.

Komentar-komentar dari para konsultan kemudian ditinjau kembali oleh staf ACT, dan perubahan yang diperlukan akan dilakukan. Jika terdapat perubahan yang besar, komponen yang direvisi kembali ditinjau oleh konsultan yang sesuai dan oleh staf ACT. Jika permasalahan kandungan tambahan muncul, konsultan yang lain diminta meninjau bentuk tersebut. Jika tidak diperlukan koreksi lebih lanjut, bentuk tes tersebut disiapkan untuk dicetak.

g) Tinjauan Setelah Pelaksanaan Tes Nasional

Segera setelah tiap pelaksanaan tes nasional, setiap bentuk tes dalam AAP dihadapkan pada suatu analisa item. Hasil dari analisa ini secara teliti diselidiki untuk mencari kejanggalan misalnya perubahan besar dari pelaksanaan pra-tes dan tes nasional dalam hal kesulitan item dan indeks diskriminasi.

Terkadang satu atau lebih peserta tes dan/atau seorang administrator tes memunculkan pertanyaan mengenai item tes tertentu setelah suatu bentuk tes dilaksanakan. Selama

tinjauan setelah pelaksanaan tes nasional, semua item yang menimbulkan masalah diteliti secara hati-hati oleh staf ACT dan ahli kandungan eksternal dan suatu respon dipersiapkan untuk pihak yang ingin mengetahui hasil tinjauan tersebut.

2) Bagian Informasi Nilai/Mata Pelajaran SMA

Mungkin penemuan riset yang paling konsisten dalam pendidikan adalah bahwa nilai SMA merupakan alat peramal nilai perguruan tinggi dan, lebih jauh, nilai tes di SMA yang digabungkan merupakan alat peramal yang lebih baik untuk nilai di perguruan tinggi dibandingkan hanya salah satu komponen tersebut. Valiga (1986) memberikan tinjauan terbaru mengenai topik ini. Reliabilitas dan validitas prediktif dari nilai SMA yang dilaporkan siswa merupakan pertimbangan penting dalam penentuan kegunaan dari alat-alat pengukur ini.

Siswa yang mendaftar untuk tanggal tes nasional diminta melaporkan nilai-nilai yang didapatkan dalam 30 mata pelajaran SMA yang berbeda dalam enam area akademik: Bahasa Inggris, matematika, IPA, IPS, bahasa, dan seni. Karena nilai SMA bergantung pada kemampuan akademik dan karakteristik pribadi seperti kerja keras dan kebiasaan studi, laporan-laporan terlapor sendiri ini memberikan perkiraan yang berguna dari prestasi akademik masa depan. Sebelum tahun akademik 1985-86, siswa yang mendaftar AAP hanya diminta untuk melaporkan nilai yang didapatkan dari kelas terakhir yang diambil dalam Matematika, IPS, Bahasa Inggris, dan IPA. Namun, kebanyakan sekolah tinggi, universitas, dan agensi negara bagian memerlukan informasi dari pendaftar mengenai performa pada sejumlah besar pelajaran SMA. Untuk memenuhi kebutuhan ini, ACT, dengan konsultasi kepada kelompok representatif dari personel lembaga-lembaga pendidikan post secondary, telah mengembangkan daftar yang lebih panjang dengan 30 mata pelajaran. Nilai siswa yang dikumpulkan oleh AAP dilaporkan pada lembaga pendidikan post secondary pilihan siswa terkait dalam Laporan Perguruan Tinggi.

3) Bagian Profil Siswa

Selain alat pengukur perkembangan pendidikan dan nilai SMA, informasi siswa lain-nya juga dikumpulkan sebagai bagian dari AAP untuk memperluas basis informasi dari siswa dan juga perguruan tinggi. Perkembangan Bagian Profil Siswa (SPS) telah dipengaruhi oleh konteks pendidikan di mana SPS berkembang, seperti juga yang dialami

bagian-bagian lain dari AAP. Asumsi utama yang mendasari perkembangan SPS adalah bahwa kualitas pendidikan yang diberikan suatu perguruan tinggi bergantung, sebagian, pada jumlah informasi yang relevan yang dipunyai stafnya mengenai siswa, dan bahwa kualitas pendidikan meningkat jika informasi ini tersedia dalam bentuk sistematis sebelum penerimaan/pendaftaran.

Bagian Profil Siswa terdiri dari beberapa sub bagian. Alasan yang mendasari perkembangan tiap sub bagian akan dibahas di bawah. Item-item SPS telah dikembangkan oleh staf ACT dengan masukan dari personel berbagai lembaga pendidikan post secondary. Item-item direvisi dari waktu ke waktu saat kebutuhan meningkat bagi lembaga-lembaga ini untuk mendapatkan tipe data yang berbeda.

Dalam bagian SPS dirancang sedemikian rupa untuk menghasilkan dua tipe informasi. Tipe pertama adalah hal yang mendasar terhadap perencanaan oleh perguruan tinggi karena mencakup rencana pendaftaran siswa (penuh waktu/paruh waktu, siang/malam, tanggal pendaftaran, jenis akomodasi kehidupan yang diinginkan, status pernikahan dan tempat tinggal). Tipe informasi kedua terkait dengan kredit perguruan tinggi sebelumnya, tugas militer, dan adanya cacat fisik atau kesulitan belajar. Instruksi SPS menyatakan dengan eksplisit bahwa informasi terakhir di atas tidak harus diungkapkan.

a) Rencana , Minat, dan Kebutuhan Pendidikan

Suatu asumsi yang mendasari perkembangan sub bagian ini adalah bahwa masuknya siswa ke dalam pendidikan post secondary mengharuskan dia untuk membuat pilihan dan keputusan tertentu, walaupun jika pemilihan tersebut bersifat sementara. Sebagai akibatnya, terjadi penyempitan pilihan pekerjaan. Faktor-faktor terkait seperti aspirasi pendidikan dan pekerjaan juga mempengaruhi keputusan siswa terhadap masa depannya.

SPS memberikan kesempatan bagia siswa untuk menunjukkan informasi-informasi seperti: jurusan perguruan tinggi yang dimaksudkan, aspirasi pendidikan gelar sarjana, IPK tahun pertama yang diperkirakan, dan rencana ekstrakurikuler. Pemberian informasi ini membantu siswa untuk mengamati rencana dan sarannya. Konselor juga diberikan data yang berguna dalam membantu siswa untuk mengevaluasi kelogisan pilihan mereka. Satu alasan lain untuk memasukan sub bagian ini adalah bahwa sub gaian ini

memberikan perguruan tinggi waktu yang lebih baik dalam merencanakan program pendidikan yang akan mereka punyai dibandingkan waktu yang mereka punyai jika data tersebut tidak diberikan sampai saat siswa mendaftar ke lembaga tersebut.

b) Kebutuhan, Minat, dan Sasaran Pendidikan Khusus

Dengan adanya kelas yang baru masuk, perguruan tinggi terkait harus siap untuk memberikan bantuan bagi kebutuhan pendidikan khusus mahasiswanya. Daftar kebutuhan tersebut mencakup penempatan tingkat lanjut dalam bidang tertentu dari kurikulum, **credit by examination**, dan bantuan dalam meningkatkan skill-skill tertentu. Dengan menyediakan informasi tersebut, siswa akan mampu menyadari perguruan tinggi terkait akan kebutuhannya. Pada saat yang sama, proses peresponan pada daftar tersebut mungkin bisa menyadarkan siswa tentang pilihan-pilihan yang sebelumnya tidak diperhatikan.

c) Rencana Ekstrakurikuler Perguruan Tinggi

Untuk membantu perguruan tinggi dalam mengembangkan program ekstrakurikuler yang tepat, informasi mengenai rencana masa datang dari siswa mereka yang akan masuk adalah hal yang berharga. Dari sudut siswa, penyajian rencana ekstrakurikuler mereka merupakan satu cara lain untuk menunjukkan pola keunikan minat, kebutuhan, dan skill mereka. Informasi yang diberikan dalam sub bagian SPS ini mencakup minat dalam organisasi, sosial, politik, dan keagamaan, dan juga seni, olah raga serta aktivitas lainnya.

Pertanyaan mengenai rencana siswa untuk membiayai pendidikan perguruan tingginya dicakup dalam sub bagian SPS ini. Informasi dari respon terhadap pertanyaan-pertanyaan ini bisa berguna bagi bagian bantuan keuangan perguruan tinggi tersebut. Siswa diminta menyebutkan suatu perkiraan penghasilan tahunan keluarga. Pertanyaan yang lainnya meminta siswa untuk menunjukkan apakah dia bermaksud untuk mengajukan permohonan bantuan keuangan dan/atau untuk bekerja paruh waktu selama kuliah.

Pertanyaan mengenai besar keluarga, besar komunitas, jarak dari perguruan tinggi, kelompok keagamaan, bahasa yang dipakai di rumah, dan latar belakang ras/etnik adalah pertanyaan-pertanyaan dalam sub bagian SPS ini. Sebagian pertanyaan ini mencakup respon pilihan, "Saya memilih untuk tidak merespon," untuk menyatakan bahwa sebagian individu mungkin memilih untuk memberikan informasi tersebut. Informasi yang

dikumpulkan dari sub bagian ini dimaksudkan untuk digunakan oleh perguruan tinggi dalam proses perencanaannya.

Informasi mengenai bagaimana siswa memilih suatu perguruan tinggi bisa menjadi berguna bagi personel yang bertanggung jawab dalam perencanaan. Sub bagian SPS ini berisi pertanyaan-pertanyaan mengenai tipe (publik/swasta, **co-educational atau tidak**, dua tahun/empat tahun), besar, lokasi, dan biaya kuliah maksimum yang diinginkan siswa dari suatu perguruan tinggi. Siswa juga diminta membuat ranking untuk faktor-faktor tersebut, bersama dengan kurikulum, urutan kepentingan keputusannya.

d) Informasi SMA

Sub bagian SPS ini meminta siswa untuk memberikan informasi mengenai tipe SMA tempat dia belajar (publik/swasta, besar, komposisi ras). Tambahan informasi diminta mengenai performa siswa sendiri (rata-rata keseluruhan, ranking) dan program SMA. Selanjutnya, siswa diminta pula untuk memilih dari suatu daftar aktivitas yang mereka ikuti di SMA. Aktivitas di daftar tersebut mewakili bidang seperti atletik, drama, musik, organisasi siswa, publikasi siswa, dan klub minat khusus.

Prestasi (penghargaan, menang pemilu siswa, produksi kreatif, dll) dalam aktivitas ekstrakurikuler saat di SMA juga mendapat perhatian. Dalam kaitannya dengan pertanyaan dalam sub bagian sebelumnya, pertanyaan-pertanyaan ini memungkinkan siswa untuk melaporkan prestasi tertentu selain juga partisipasi dalam rentang luas aktivitas di luar kelas. Sub bagian SPS ini juga memberikan kesempatan pada siswa untuk mengungkapkan tingkat kepuasan dia terhadap beragam aspek SMA: instruksi/pengajaran, beragam mata pelajaran, fasilitas, layanan dan program lain. Juga diminta suatu penilaian keseluruhan kelayakan pendidikan SMA dalam skala dengan lima angka (sangat tidak layak sampai sangat baik).

4) Inventori Minat ACT

Holland (1973, 1985), Roe (1956) dan yang lainnya telah menyatakan bahwa domain minat pekerjaan bisa diwakili oleh sedikit tipe minat dasar. Skala inventori minat ACT dikembangkan ada enam tipe sesuai dengan yang dikembangkan Holland.

Beberapa inventori minat yang berbeda namun terkait telah dikembangkan dan digunakan dalam beragam program ACT. Bentuk yang sekarang digunakan adalah Edisi Unisex dari Daftar Minat ACT (UNIACT). UNIACT disusun dengan tujuan agar

distribusi pilihan karir yang sama akan ditunjukkan pada pria maupun wanita. Dalam konteks ini, item-item di mana pria dan wanita mempunyai distribusi respon yang sama disebut "**item sex-balanced.**"

Dalam perkembangan UNIACT, item-item yang ditulis untuk menangkap esensi beragam aktivitas yang terkait dengan kerja ditambahkan pada suatu pool item dari daftar minat ACT sebelumnya, sehingga menghasilkan lebih dari 200 kemungkinan item sex-balanced. Analisa item dilaksanakan pada enam sampel, melibatkan lebih dari 10.000 siswa kelas 9 dan 11, siswa yang ingin melanjutkan ke perguruan tinggi, dan orang dewasa (lihat **bab 3** dari laporan teknis UNIACT [ACT, 1981a] untuk rincian lebih lanjut). Kebanyakan sampel juga diberi satu bentuk lain Daftar Minat ACT, dan korelasi antara item UNIACT yang akan digunakan dan skala pada daftar-daftar ini lalu dihitung. Pada setiap tahap pengembangan skala, penseleksian dan perbaikan item melibatkan pengamatan persentase respon "suka" untuk pria dan wanita, dan korelasi item dengan skala pada daftar sebelumnya dan juga skala UNIACT yang akan digunakan. Dengan cara ini, keenam skala UNIACT dikembangkan sehingga domain kandungan yang diwakili oleh item tersebut secara langsung terkait dengan teori Holland dan item-item tersebut menjadi sex-balanced.

Informasi rinci mengenai dasar pemikiran UNIACT, perkembangan, dan karakteristik psikometrik terdapat pada *Technical Report for the Unisex Edition of the ACT Interest Inventory (UNIACT)* (ACT, 1981a). Laporan ini juga mencakup profil minat untuk lebih dari 40.000 orang dalam 352 kelompok pekerjaan dan pendidikan.

Pernyataan-pernyataan berikut menjelaskan bidang yang dicakup oleh keenam skala minat tersebut. Tipe-tipe minat Holland yang terkait ditunjukkan dalam tanda kurung setelah nama skala.

Sains (**Investigative**) Menyelidiki dan mencoba memahami fenomena dalam IPA melalui proses membaca, riset, dan diskusi.

Seni (**Artistic**) Mengekspresikan diri melalui aktivitas seperti melukis, mendesain, menyanyi, menari, dan menulis; apresiasi artistik dari aktivitas-aktivitas tersebut (misal, menyimak musik, membaca karya sastra).

Layanan Sosial (**Social**) Membantu, meringankan, atau melayani orang lain melalui aktivitas-aktivitas seperti mengajar, konseling, bekerja pada organisasi berorientasi layanan/jasa, dan terlibat dalam studi sosial/politik.

Kontak Bisnis (**Enterprising**) membujuk, mempengaruhi, mengarahkan, atau memotivasi orang lain melalui aktivitas-aktivitas seperti penjualan, pengawasan, dan aspek-aspek manajemen bisnis.

Operasi Bisnis (**Conventional**) Mengembangkan dan/atau memelihara kerapian dan ketepatan file, catatan, rekening, dll; merancang dan/atau mengikuti prosedur sistematis untuk melaksanakan aktivitas bisnis.

Teknis (**Realistic**), bekerja dengan alat, instrumen, dan perlengkapan mesin atau elektronik. Aktivitas-aktivitasnya mencakup merancang, membangun, dan memperbaiki mesin dan memelihara tanaman/ternak.

Pengembangan Peta Dunia Kerja, yang dirancang untuk membantu siswa dan konselornya dalam menafsirkan hasil UNIACT, dijelaskan di bawah ini.

a) Peta Dunia Kerja

Karena terdapat banyak sekali jenis pekerjaan - lebih dari 12.000 yang terdaftar dalam edisi keempat *Dictionary of Occupational Titles* (United States Department of Labor, 1977) - ACT telah mengembangkan suatu sistem pengelompokan pekerjaan yang membuat eksplorasi pekerjaan menjadi lebih mudah bagi para siswa. Pengelompokan pekerjaan ACT didasarkan pada campuran empat tugas kerja dasar: Data, Ide, Orang, dan Benda dari setiap pekerjaan.

Banyak konselor yang tidak asing dengan model hexagonal dari Holland untuk enam tipe minat dan pekerjaan. Seperti dicatat di muka, laporan SMA dan laporan perguruan tinggi memasukkan skor tiap siswa pada enam skala minat UNIACT, yang serupa dengan hal ini. Keenam skala ini terkait dengan Dimensi tugas kerja Data-Ide dan Orang-Benda seperti ditunjukkan pada Gambar 1 (ACT, 10981a; Prediger, 10981, 1982). Perhatikan bahwa kedua dimensi tugas kerja merangkum dua dimensi yang mendasari keenam skala. Seseorang yang mendapat skor tertinggi pada skala Layanan Sosial kemungkinan akan memilih tugas-tugas kerja Orang; orang dengan skor paling tinggi pada skala Kontak Bisnis kemungkinan akan memilih tugas kerja Data dan Orang.

Gambar 1. Dimensi Data/Ide dan Orang/Benda

Sistem ACT untuk pengelompokan pekerjaan dirangkum oleh Daftar Keluarga Karir yang dijelaskan dalam *Using the ACT Assessment on Campus* (ACT, 1987i) dan *The ACT Assessment Counselor's Handbook* (ACT, 1987j). Pekerjaan di dalam tiap keluarga karir relatif serupa dengan melihat pada keterlibatan tugas kerja Data-Ide dan Orang-Benda, tujuan kerja, dan setting kerja (Prediger, 1976). keluarga-keluarga di dalam daftar diorganisasikan ke dalam enam kelompok yang berkaitan dengan keenam skala UNIACT. Peta Dunia Kerja menunjukkan di mana letak tiap keluarga karir pada dimensi tugas kerja. Contohnya, Keluarga karir B (manajemen dan Perencanaan) terletak pada wilayah 3, yang menunjukkan keterlibatan dengan Orang dan Data. Yang digunakan untuk mengembangkan edisi kedua Peta Dunia Kerja adalah rating analisa kerja Departemen

Buruh Amerika untuk kira-kira 13.000 pekerjaan tersendiri dalam *Dictionary of Occupational Titles* (United States Department of Labor, 1977) dan juga skor minat untuk 421 kelompok pekerjaan dan pendidikan.

Para siswa memerlukan suatu cara untuk menerjemahkan skor UNIACT mereka ke dalam pekerjaan yang sesuai dengan minat mereka. Untuk tujuan ini, hasil UNIACT dilaporkan pada Laporan Siswa dan Laporan SMA sebagai wilayah peta pada Peta Dunia Kerja. Alasan dan metode untuk penghitungan koordinat siswa pada peta tersebut dijelaskan secara rinci oleh Prediger (1981). Berdasar wilayah peta siswa, Laporan Siswa AAP menyarankan rumpun karir yang mungkin ingin diketahui oleh siswa. Rumpun karir ini berada dalam atau bersebelahan dengan wilayah peta siswa terkait. Contoh-contoh pekerjaan dalam tiap rumpun karir, jurusan perguruan tinggi yang terkait, dan acuan pada deskripsi pekerjaan diberikan pada Laporan Siswa. Peta Dunia Kerja ditunjukkan dan dijelaskan dalam *Using Your ACT Assessment Result* (ACT, 1987m), yang didistribusikan pada semua peserta tes dengan hasil AAP mereka.

Klasifikasi pekerjaan ke dalam rumpun karir Peta Dunia Kerja dan lokasi rumpun kerja pada peta tersebut didasarkan pada suatu sintesa dari beragam data empiris. Seperti dijelaskan oleh Prediger (1976), data-data ini mencakup karakteristik pekerjaan tertentu dalam *Dictionary of Occupational Titles* dan profil daftar minat dari kelompok kerja yang telah mengambil **Blank** Minat Kuat Pekerjaan, skala minat Proyeksi Bakat, atau Daftar Pilihan Kerja dari Holland. Profil minat tersedia untuk sekitar 110.000 orang dalam 570 kelompok kerja.

b) Keadilan terhadap Jenis Kelamin dalam UNIACT

Garis panduan National Institute of Education mengenai keadilan terhadap jenis kelamin dalam inventori minat (Diamond, 1975) menyatakan bahwa "Minat dan pilihan pekerjaan dari pria dan wanita dipengaruhi oleh banyak faktor lingkungan dan budaya". Satu faktor, pen-stereotipe-an peran jenis kelamin, terkadang mempengaruhi respon siswa pada item-item inventori minat. Sebagai hasilnya, kecuali jika dilakukan tindakan yang hati-hati, saran karir yang disediakan oleh suatu inventori minat bisa terbatas ke dalam bidang-bidang secara tradisi dipegang oleh jenis kelamin mereka. Walaupun keseimbangan jenis kelamin yang sempurna belum dicapai pada skala UNIACT, pilihan item yang hati-hati telah menghasilkan skala yang cukup seimbang tanpa mengorbankan

reliabilitas dan validitas (ACT, 1981a). Pria dan wanita mendapatkan skor standar yang serupa pada skala UNIACT.

c. Bukti Validitas untuk Instrumen Penilaian ACT

Menurut Brown (1980), "jika satu karakteristik dari suatu tes bisa dianggap lebih penting dari yang lainnya, karakteristik itu adalah validitas". Namun validitas bukan satu-satunya karakteristik suatu instrumen. Seperti dinyatakan dalam *Standards for Educational and Psychological Testing* (AERA, APA, dan NCME, 1985), "Konsep validitas mengacu pada ketepatan, kebermaknaan, dan kebergunaan dari kesimpulan tertentu yang ditarik dari skor tes". Lebih jauh, walaupun terdapat tiga kategori bukti validitas yang umumnya diakui, "perbedaan yang bertolak belakang antar ketiga kategori tersebut tidaklah dimungkinkan". Contohnya, bukti yang biasanya teridentifikasi dengan kategori kriteria atau isi juga relevan dengan kategori konstruk". Untuk mudahnya, dalam kesempatan ini akan dibahas **Tes Perkembangan Pendidikan** yang merupakan komponen dari program asesmen ACT.

Tujuan utama tes perkembangan pendidikan adalah memberikan informasi yang membantu siswa, orang tua, guru, konselor, dan petugas penerimaan perguruan tinggi untuk membuat keputusan logis dengan menimbang proses transisi dari pendidikan SMA (secondary) ke pendidikan post secondary. Berikut akan dibahas secara sepintas validitas isi, kriteria, dan validitas konstruk.

1) Validitas Isi

Ide fundamental yang mendasari perkembangan keempat tes tersebut adalah bahwa cara terbaik untuk memperkirakan keberhasilan di perguruan tinggi adalah dengan mengukur selangsiung mungkin kemampuan yang harus diterapkan siswa dalam tugas kuliahnya. Hal ini berarti tugas-tugas yang diberikan dalam tes harus mewakili tugas perkuliahan/sekolah. Tugas tersebut harus rumit dalam strukturnya; harus komprehensif dalam cakupannya; dan harus signifikan bukannya tugas yang sempit atau dibuat-buat yang dimasukan dengan alasan korelasi statistiknya dengan suatu kriteria.

Tes AAP berisi suatu proporsi besar latihan pemecahan masalah yang rumit dan beberapa alat pengukur skill sempit yang jumlahnya proporsional. Tes tersebut diorientasi-kan pada bidang utama dari program pengajaran perguruan tinggi dan SMA bukannya pada suatu definisi faktorial dari beragam aspek kecerdasan. Skor-skor tersebut

mempunyai hubungan langsung dan jelas dengan kemajuan pendidikan siswa, dan suatu makna yang dapat segera ditangkap baik oleh staf pengajar maupun siswa terkait.

Karena tes-tes tersebut dirancang untuk mengukur sejauh mana siswa telah mengem-bangkan skill dan kemampuan umum yang diperlukan untuk keberhasilan dalam tugas kuliah, *content-validity* sangatlah penting.

Seperti dijelaskan di muka, prosedur perkembangan tes AAP mencakup suatu proses peninjauan yang ekstensif, yang setiap itemnya diamati secara kritis setidaknya dua belas kali. Spesifikasi tes yang rinci telah dikembangkan untuk menjamin bahwa kandungan tes tersebut mewakili kurikulum universitas dan SMA yang sedang berlaku. Semua bentuk ditinjau untuk menjamin bahwa bentuk tersebut memenuhi spesifikasi-spesifikasi ini. Satu komponen utama proses peninjauan adalah tinjauan terhadap isi atau kandungan test oleh konsultan luar yang merupakan ahli di bidang yang sedang dinilai. Karena itu, terdapat suatu penilaian yang berkelanjutan untuk validitas isi dari tes tersebut selama prose pengembangannya. Tes sampel (bentuk yang di"release" yang digunakan dalam tahun-tahun sebelumnya) bisa didapatkan dari ACT untuk pihak yang ingin mempelajari kandungan test tersebut.

Dalam analisa Test Penggunaan Matematika oleh model teori "generalizability", Jarjoura dan Brennan (1982, 1983) mencatat bahwa test tersebut sangat dekat kesesuaiannya dengan tabel spesifikasi. Suatu analisa distribusi skor universal untuk tiap bidang kandungan (operasi aritmetika dan aljabar, reasoning aritmetika dan aljabar, geometri, aljabar tingkat menengah, konsep angka dan numerasi, dan topik lanjutan lain) terhadap varian universal komposit menghasilkan penemuan bahwa kontribusi proporsional dari keenam kategori (.11, .34, .20, .20, .14) sangatlah dekat dengan proporsi item yang ditentukan dalam tabel spesifikasi (.10, .35, .20, .20, .15). Hasil ini menunjukkan bahwa kepentingan relatif berdasar teori dari bidang kandungan diwakili dengan baik dalam bentuk Test Penggunaan Matematika yang senyatanya.

2) Validitas Kriteria Terkait

Hakikat/esensi dari validitas ini adalah hubungan antara skor test dengan alat pengukur kriteria luar. Dalam validitas kriteria terkait ini, validitas sering dibedakan menjadi dua, yaitu: validitas prediktif, dan validitas konkuren. Seperti dinyatakan dalam *Standards for Educational and Psychological Tests* (AERA, APA, dan NCME, 1985),

"Suatu studi prediktif berisi informasi mengenai keakuratan yang dapat digunakan data tes awal untuk memperkirakan skor kriteria yang akan didapatkan di masa datang. Studi konkuren berfungsi untuk tujuan yang sama, tapi studi ini mendapatkan informasi prediksi dan kriteria pada saat yang sama".

Skor tes ACT digunakan oleh lembaga pendidikan post secondary umumnya untuk membantu mereka dalam membuat dua tipe keputusan: keputusan pemilihan dan keputusan penempatan. Kriteria penting untuk mengevaluasi skor pada tes perkembangan pendidikan adalah nilai universitas secara keseluruhan dan nilai dalam mata kuliah tertentu. Seperti yang bisa diperkirakan, kebanyakan bukti validitas yang diberikan dalam bab ini berkaitan dengan tingkat hubungan antara skor tes dengan nilai di universitas.

Informasi Validitas Prediktif dari Laporan Layanan Riset. Divisi riset ACT secara rutin memberikan data validitas prediktif melalui dua layanan riset yang dibahas di muka. Layanan Riset Dasar menghasilkan laporan tahunan yang menjelaskan prestasi akademik mahasiswa pertama saat itu, dan mengembangkan persamaan untuk memperkirakan IPK semester pertama untuk mahasiswa tahun mendatang. Laporan Riset Dasar ini dapat didapatkan oleh lembaga-lembaga dengan setidaknya 50 siswa yang pernah dites AAP. Untuk mendapatkan laporan tersebut, suatu lembaga harus memberi ACT catatan akademik dari setidaknya sejumlah siswa tersebut. Jika catatan setidaknya 50 siswa dari tiap jenis kelamin sudah didapatkan, lembaga tersebut akan menerima laporan terpisah sesuai jenis kelamin dan juga laporan keseluruhan. Informasi mengenai partisipasi dalam layanan ini bisa didapatkan dalam publikasi ACT *Instructional Guide for the ACT basic Research* (ACT, 1984). Suatu deskripsi rinci mengenai bagaimana menggunakan laporan ini diberikan dalam publikasi yang lain, *Your College Freshmen: Interpretive Guide to ACT Research Services for Higher Education* (ACT, 1981b).

Laporan Layanan Riset Dasar memberikan interkorelasi antara empat skor tes AAP, skor Composite AAP, nilai rata-rata SMA, dan IPK perguruan tinggi untuk siswa dari suatu universitas. Laporan ini juga memberikan koefisien korelasi ganda IPK dengan keempat tes AAP ; dan koefisien korelasi ganda IPK dengan keempat tes AAP dan nilai rata-rata SMA, serta juga rata-rata (means) dan standar deviasi dari semua variabel. Laporan Riset Dasar ini meringkas keakuratan prediktif pada semua perguruan tinggi

yang menjadi partisipan dalam Layanan Riset Dasar pada tahun sebelum tahun pelaporan tersebut.

Ditemukan bahwa untuk 510 perguruan tinggi yang termasuk dalam laporan, korelasi ganda median untuk memperkirakan IPK semester pertama perguruan tinggi dari keempat tes AAP adalah 0,48 untuk kelas mahasiswa tahun pertama 1985-86. Juga karena kuartil pertama dan ketiga dari distribusi korelasi ganda tersebut adalah 0,39 dan 0,56, setengah dari perguruan tinggi yang berpartisipasi mempunyai nilai di antara kedua nilai tersebut. Sebagai perbandingan, korelasi median antara rata-rata nilai SMA dan IPK tahun pertama perguruan tinggi adalah 0,44, dengan titik kuartil pertama dan ketiga adalah 0,36 dan 0,53. Jika nilai rata-rata SMA dimasukkan bersama keempat skor tes AAP dalam suatu regresi ganda dengan lima alat prediksi, korelasi ganda median adalah 0,55, dengan 0,47 dan 0,63 sebagai titik kuartil pertama dan ketiga dalam distribusi tersebut. Standar kesalahan median untuk perkiraan tersebut adalah 0,66 unit nilai, dengan kuartil pertama dan ketiga pada distribusi tersebut adalah 0,58 dan 0,76 unit nilai. Median standar deviasi kesalahan estimasi diperkirakan sebesar 0,66 dari IPK semester pertama, dan setengah perguruan tinggi yang berpartisipasi mempunyai standar kesalahan estimasi tersebut antara 0,58 dan 0,76.

Layanan riset prediksi kedua yang diberikan oleh ACT adalah Layanan Riset Standar. Seperti juga Layanan Riset Dasar, layanan menjelaskan prestasi akademik dan potensi mahasiswa tahun pertama yang masuk dan mengembangkan persamaan prediksi yang digunakan dalam memperkirakan performa keseluruhan siswa di masa datang. Melalui Layanan Standar suatu perguruan tinggi bisa mengembangkan persamaan prediksi untuk nilai dengan jumlah mata kuliah sampai 40 buah, dan sub kelompok terpisah siswa sampai sembilan buah. Layanan ini juga memungkinkan perguruan tinggi untuk mengembangkan persamaan prediksi dengan menggunakan alat prediksi yang dapat didapatkan secara lokal, seperti ranking sekolah yang tinggi, skor tes penempatan lokal, dan skor SAT. Layanan Riset Standar ini bisa didapatkan oleh semua lembaga yang menggunakan AAP dan mempunyai setidaknya 50 catatan siswa AAP untuk tiap sub kelompok yang akan dimasukkan. Untuk berpartisipasi dalam layanan ini, perguruan tinggi terkait harus menyediakan semua data yang diperlukan mengenai performa mahasiswanya pada semester atau tahun pertama mereka di perguruan tinggi tersebut.

ACT merekomendasikan bahwa perguruan tinggi yang berpartisipasi dalam layanan ini mempunyai akses pada seorang ahli statistik atau peneliti lembaga lain yang berkualitas. Informasi spesifik mengenai Layanan Riset Standar diberikan dalam *Prosedures for Participating in the ACT Standard Research Service* (ACT, 1987d), dan informasi mengenai penggunaan laporan tersebut dalam *Your College Freshmen: Interpretive Guide to ACT Research Services for Higher Education* (ACT, 1981b).

3) Validitas Konstruk

Validitas konstruk adalah validitas yang didasarkan pada logika atau teori yang dikaji. Suatu instrumen dikatakan mempunyai validitas konstruk yang baik manakala alat ukur itu secara logika mampu mengukur apa yang seharusnya diukur. Alat ukur berisi butir-butir yang menurut teori merupakan pecahan atribut yang akan diukur. Dapat pula dikatakan suatu instrumen dikatakan mempunyai validitas konstruk yang baik manakala hasil pengambilan data dengan instrumen itu sesuai dengan yang diprediksikan secara teori. Namun sekali lagi bahwa hasil statistik ini bukan segalanya, dan yang penting adalah kebermaknaan atau esensi dari butir-butir tes yang ada di instrumen tersebut.

d. Reliabilitas dari Instrumen Penilaian ACT

Estimasi reliabilitas dan standar kesalahan pengukuran untuk tiap tes AAP dilaporkan dalam bagian ini. Sub-sub bagian di sisni secara terpisah berkaitan dengan konsistensi internal dan estimasi kestabilan.

1) Perkiraan Konsistensi Internal

Setelah tanggal tes nasional, suatu sampel acak dari 2.000 peserta tes dipilih dan suatu analisa item dilakukan pada keempat tes untuk para peserta tes ini. Formula Kuder-Richardson 20 digunakan untuk mengestimasi reliabilitas dan standar kesalahan pengukuran yang menyertainya dihitung sebagai bagian dari analisa item ini.

Karena koefisien KR20 dihitung dari statistik item, koefisien tersebut mewakili estimasi reliabilitas skor mentah dari keempat tes. Karena tidak mungkin menghitung koefisien seperti itu untuk mencari skor standar, koefisien KR20 ini digunakan sebagai estimasi reliabilitas dari skor standar. Adalah suatu kemungkinan bahwa koefisien skor mentah KR20 merupakan perkiraan berlebihan untuk reliabilitas sebenarnya dari skor standar keempat tes AAP. Sejauh mana koefisien itu melebih-lebihkan perkiraan reliabilitas skor standar belum dapat ditentukan. Namun, karena transformasi skor mentah

menjadi standar mempertahankan urutan ranking peserta tes, maka perbedaan dalam reliabilitas untuk skor standar dan mentah kemungkinannya hanyalah tipis.

Standar kesalahan dari pengukuran untuk mencari skor standar bisa diperhitungkan dengan mengkalikan standar deviasi skor tersebut dengan akar kuadrat satu minus perkiraan reliabilitas.