

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Dalam rangka ikut mencerdaskan kehidupan bangsa dan memberi kesempatan masyarakat untuk mengenyam pendidikan tinggi yang berkualitas, di setiap tahunnya Universitas Negeri Yogyakarta menerima mahasiswa baru yang jumlahnya cukup banyak. Penerimaan mahasiswa baru tersebut melalui tiga macam cara, yaitu : penelusuran bibit unggul (PBU), seleksi penerimaan mahasiswa baru (SPMB) dengan ujian tertulis, dan seleksi yang dilakukakn oleh UNY sendiri. PBU dan SPMB dilaksanakan dengan tujuan untuk menerima mahasiswa program reguler. Khusus untuk SPMB, pelaksanaannya dilaksanakan secara nasional bersama-sama dengan PTN lain dan dapat diikuti oleh calon mahasiswa secara lintas daerah/wilayah. Seleksi yang dilakukan oleh UNY sendiri meliputi tiga macam, yaitu : seleksi untuk masuk program D3 Reguler Fakultas Teknik dengan Ujian Tertulis, Penelusuran Bibit Daerah (PBD), dan seleksi dengan ujian tertulis untuk masuk program S1 dan D3 Non Reguler di lingkungan UNY.

Berdasarkan informasi yang diperoleh sebelumnya, diperkirakan bahwa jumlah animo calon yang masuk program Non Reguler UNY Program D3 dan S1 Sore ini jauh lebih banyak daripada daya tampungnya. Daya tampung yang ada hanya sekitar 600 dari sekitar 6500 calon mahasiswa. Oleh karena daya tampung jauh lebih kecil daripada animo masuk, maka alat seleksi masuk Program D3 dan S1 Sore ini perlu disiapkan dengan sebaik-baiknya. hal ini dilakukan agar dalam penyelenggaraan seleksi tersebut tidak ada pihak-pihak yang dirugikan. Alat seleksi tersebut berupa suatu tes tertulis yang harus diikuti oleh setiap calon mahasiswa baru.

Secara garis besar, jenis tes dapat dikelompokkan menjadi lima, yaitu : (1) Binet dan Tes Intelegensi, (2) Testing Kelompok, (3) Tes Potensi Intelektual atau Tes Potensi Akademik, (4) Tes Hasil Belajar, dan (5) Tes Projektif. Untuk seleksi masuk UNY Program D3 dan S1 Non Reguler UNY ini adalah Tes Potensi Akademik (TPA). Dalam hal ini, instrumen terdiri dari empat subtes, yaitu: subtes Kemampuan Verbal, Kemampuan Kuantitatif, Kemampuan Penalaran, dan subtes Kemampuan Bahasa Inggris. Subtes Kemampuan verbal terdiri dari 30 butir soal, Kemampuan

Kuantitatif terdiri dari 30 butir soal, Penalaran terdiri dari 35 butir soal, dan subtes Bahasa Inggris terdiri dari 15 butir soal.

Sebagaimana telah disampaikan di atas, agar tidak ada pihak-pihak yang dirugikan dalam penerimaan mahasiswa baru, perlu digunakan instrumen atau alat seleksi yang berupa soal tes tertulis yang baik. Dengan demikian perlu diketahui bagaimana kualitas dari soal tes yang digunakan dalam seleksi masuk program S1 dan D3 Non Reguler UNY. Oleh karena itu, penelitian ini perlu dilakukan agar dapat diperoleh informasi yang tepat tentang karakteristik butir soal ujian masuk program S1 dan D3 Non Reguler tahun 2005.

B. Rumusan Masalah

Berdasarkan uraian di atas, maka masalah dalam penelitian ini dapat dirumuskan sebagai berikut.

Bagaimanakah karakteristik butir soal ujian masuk program S1 dan D3 Nonreguler UNY tahun 2005, apabila ditinjau dari aspek : indeks kesukaran, daya beda, peluang tebakan, validitas dan reliabilitasnya dan hubungan antar sub tesnya.

C. Tujuan Penelitian

Tujuan pokok penelitian ini adalah ingin mengetahui karakteristik butir soal yang telah disusun dalam rangka ujian masuk program S1 dan D3 Nonreguler UNY tahun 2005 , apabila ditinjau dari aspek : indeks kesukaran, daya beda, peluang tebakan, validitas dan reliabilitasnya dan mengetahui hubungan antar sub tes pada tes potensi akademik untuk calon mahasiswa UNY non reguler.

D. Manfaat Penelitian

Manfaat hasil penelitian ini adalah sebagai bahan informasi bagi khususnya Pusat Studi Pengembangan Sistem Pengujian Lembaga Penelitian UNY sebagai penyusun perangkat tes dan Panitia Penerimaan Mahasiswa Baru tentang kualitas soal

yang telah digunakan dalam pelaksanaan seleksi mahasiswa baru. Dengan demikian hasil penelitian ini diharapkan dapat dijadikan sebagai bahan pertimbangan dalam penyusunan dan pengembangan tes ujian masuk UNY sehingga dihasilkan tes yang benar-benar sah (*valid*) dan konsisten (*reliable*). Dengan kata lain, dengan adanya penelitian ini akan diperoleh soal tes masuk UNY nonreguler yang berkualitas tinggi.

BAB II

KAJIAN PUSTAKA

A. Jenis-jenis Tes

Dalam evaluasi pendidikan, diperlukan alat (instrumen). Alat yang digunakan untuk melakukan evaluasi, salah satunya adalah tes. Tes ini digunakan untuk mengetahui informasi tentang aspek psikologis tertentu. Menurut Cronbach (Nurkholis, 2000: 14), tes merupakan suatu prosedur sistematis untuk mengamati dan menggambarkan satu atau lebih karakteristik seseorang dengan suatu skala numerik atau sistem kategorik. Berdasarkan hal ini, tes memberikan informasi yang bersifat kualitatif dan kuantitatif.

Tes dapat diklasifikasikan dengan beberapa macam, tergantung dari tujuannya (Anastasi dan Urbina, 1997 : 2-4). Tes prestasi belajar merupakan suatu bentuk tes untuk mendapatkan data, yang merupakan informasi untuk melihat seberapa banyak pengetahuan yang telah dimiliki dan dikuasai oleh seseorang sebagai akibat dari pendidikan dan pelatihan (Anastasi dan Urbina, 1997: 42-43). Berdasarkan informasi yang diperoleh ini, pada proses seleksi, siswa dapat dikelompokkan sesuai dengan kemampuannya, yang diterima atau tidak diterima. Hal ini sesuai dengan fungsi tes prestasi seperti yang dikemukakan Gronlund (1976: 16), yang menyatakan bahwa tes prestasi berfungsi sebagai alat untuk penempatan, fungsi formatif, fungsi diagnostik dan fungsi sumatif.

Berdasarkan bentuknya, tes prestasi belajar dapat dikelompokkan menjadi tiga jenis, yaitu : 1) objektif, yang sederhana terdiri dari bentuk jawaban singkat, benar-salah atau dua pilihan, dan menjodohkan, serta objektif pilihan ganda dengan alternatif jawaban lebih dari dua, 2) uraian (Gronlund, 1976: 144). Mehren dan Lehmann (1973), menggolongkan tes menjadi tiga bagian, yaitu : 1) tes bakat, 2) tes prestasi, dan 3) tes minat, kepribadian, dan sikap. Menurut Suryabrata (2001), paling tidak ada lima jenis tes, yaitu : (1) Binet dan Tes Intelegensi, (2) Testing Kelompok, (3) Tes Potensi Intelektual atau Tes Potensi Akademik, (4) Tes Hasil Belajar, dan (5) Tes Projektif. Di samping itu, Azwar (1987) mengemukakan bahwa tes diklasifikasikan menjadi empat golongan, yaitu : 1) tes intelegensi umum, 2) tes

kemampuan khusus, 3) tes prestasi, dan 4) tes kepribadian. Pada tes seleksi masuk UNY program S1 dan D3 non reguler termasuk dalam Tes Potensi Akademik.

B. Alat Ukur dan Hasil Pengukuran

Pada prinsipnya, soal tes atau alat ukur lainnya digunakan untuk mengukur. Menurut Ebel (1972) pengukuran adalah pemberian angka pada seseorang atau sesuatu obyek yang dimaksudkan untuk membedakan tingkat orang atau obyek itu mengenai hal (trait) yang diukur.

Sementara itu, menurut tulisan Delandshere dan Petrosky (1998) menjelaskan bahwa 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. Nunnally (1978) juga 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.

Lebih rinci lagi, Suryabrata (2000) menjelaskan bahwa pengukuran pada hakekatnya adalah kuantifikasi. Hasilnya berupa bilangan yang menunjukkan besaran atribut yang diukur. Dengan hasil seperti itu maka akan diperoleh keuntungan-keuntungan sebagai berikut :

- a. Memungkinkan orang menyajikan data secara rinci dan menyeluruh/lengkap.

- b. Memungkinkan penerapan metode analisis matematis yang dalam ilmu pengetahuan diakui sangat kuat (*powerfully*).
- c. Memungkinkan mengembangkan teori dalam model matematis.
- d. Tingginya komunikabilitas hasil yang diperoleh sehingga prinsip ilmu pengetahuan sebagai usaha terbuka (*public enterprise*) benar-benar dapat diwujudkan.

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.

Pengukuran itu sendiri dapat dilakukan melalui tes dan dapat pula tidak melalui tes. Menurut para ahli yang dirangkum oleh Kartowagiran (1995) testing adalah suatu proses kuantifikasi potensi kognitif dan psikomotor yang menggunakan alat yang dirancang secara khusus. Ujian bagi siswa merupakan proses kuantifikasi prestasi belajar siswa dalam bidang tertentu yang melalui tes. Mengukur panjang meja atau tinggi badan seseorang merupakan kuantifikasi suatu obyek yang tidak melalui tes.

Anastasi (1976), berpendapat bahwa tes pada dasarnya merupakan suatu pengukuran yang objektif dan standar terhadap sample perilaku. Brown (1976), mengatakan bahwa tes adalah prosedur yang sistematis guna mengukur sampel perilaku seseorang. Sejalan dengan ahli lainnya, Cronbach (1970) mengatakan bahwa tes adalah prosedur yang sistematis untuk mengobservasi perilaku seseorang dan mendeskripsikan perilaku itu dengan skala numerik atau sistem kategori.

Sementara itu, Umar (2000) berpendapat bahwa testing adalah pengukuran karakteristik seseorang, khususnya kemampuan atau penge-tahuannya di bidang tertentu, dengan cara menyajikan stimulus (butir soal), yang kemudian direspons oleh orang tersebut. Atas jawaban (respons) yang diperoleh itu lalu ditetapkan kedudukan orang tersebut pada suatu skala ukuran yang ditentukan. Kedudukan atau ukuran (nilai berupa angka) yang diperoleh dapat dibandingkan dengan suatu standar (nilai

angka tertentu) pada skala yang sama sehingga dapat diambil keputusan tentang orang tersebut (misalnya, kelulusan). Kadang-kadang diperlukan pengukuran pada lebih dari satu jenis pengetahuan/kemampuan dalam penentuan kelulusan tersebut, namun tetap saja sejumlah ukuran (angka) tersebut harus digabungkan terlebih dahulu menjadi satu dimensi gabungan (*composite*). Sering kali diperlukan pembobotan dalam penggabungan ini. Ujian akhir adalah salah satu jenis dari kegiatan testing, dan kegunaan yang utama adalah untuk mengambil keputusan tentang orang yang diuji (misalnya untuk keperluan sertifikasi/kelulusan, seleksi, penjurusan, dan sebagainya).

Pengacuan hasil tes calon mahasiswa perlu dilakukan sebab nilai mentah yang diperoleh calon mahasiswa belum dapat diinterpretasikan secara tepat tanpa adanya suatu acuan. Sebagai contoh, ada calon mahasiswa yang mendapat skor tes 70 dalam ujiannya. Informasi ini belum menjelaskan kinerja siswa tersebut sebelum diketahui berapa persen siswa yang mendapat skor 70. Atau, sebelum diketahui simpang baku dan skor rerata yang diraih oleh populasi calon mahasiswa yang mengikuti ujian. Pendekatan ini biasanya disebut dengan Penilaian Acuan Norma (PAN), dan cara inilah yang biasanya digunakan dalam seleksi mahasiswa baru.

C. Kualitas Soal

Suatu alat ukur atau soal tes dikatakan berkualitas tinggi manakala soal tes itu mampu mengukur secara tepat dan hasil pengukurannya dapat dihandalkan; soal tes itu valid dan reliabel. Suatu tes dapat dikatakan mempunyai validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang tepat dan akurat sesuai dengan maksud dikenakannya tes tersebut (Azwar, 1996). Sisi lain yang sangat penting dalam konsep validitas adalah kecermatan pengukuran, yakni kemampuan untuk mendeteksi perbedaan-perbedaan kecil sekalipun yang ada pada atribut yang diukurnya.

Pengertian validitas sangat erat kaitannya dengan tujuan pengukuran. Tidak ada validitas yang berlaku secara umum untuk semua tujuan pengukuran. Suatu tes hanya valid untuk satu tujuan pengukuran yang spesifik. Sebagai contoh, tes matematika mungkin valid sebagai prediktor ilmu pengetahuan atau *science* tetapi mungkin kurang valid sebagai prediktor sejarah.

Secara empiris suatu tes adalah valid jika memenuhi dua kriteria, yaitu: (1) tes tersebut mengukur konsep atau variabel yang diharapkan hendak diukur dan tidak mengukur konsep atau variabel lain yang tidak diharapkan untuk diukur, dan (2) tes tersebut dapat memprediksi perilaku yang lain yang berhubungan dengan variabel yang diukur.

Ada berbagai type validitas, tiga type utama validitas yang erat kaitannya dengan pengukuran pendidikan dan psikologi yaitu validitas isi, validitas berdasarkan kriteria, dan validitas konstruk (Nunnally, 1978, Allen & Yen, 1979, Fernandes, 1984 dan Woolfolk & McCane, 1984).

Validitas isi menunjuk seberapa jauh tes mengukur keseluruhan kawasan pokok bahasan dan perilaku yang hendak diukur. Pengertian mencakup keseluruhan kawasan isi tidak saja berarti tes itu harus komprehensif, akan tetapi isinya harus tetap relevan dan tidak keluar dari batasan tujuan pengukuran. Validitas berdasarkan kriteria menunjukkan bahwa skor tes secara sistematis berhubungan dengan satu atau lebih kriteria hasil. Validitas berdasarkan kriteria dibedakan menjadi dua, yaitu validitas prediktif dan validitas konkuren. Menurut Fernandes (1984), validitas berdasarkan kriteria menjawab pertanyaan: *“How well test performance predicts future performance (predictive validity) or estimate current performance on some valued measure other than the test itself (concurrent validity)?”*

Validitas konstruk menunjuk pada seberapa jauh suatu tes mengukur *trait* atau konstruk teoritik yang hendak diukurnya. Untuk pengujian validitas konstruk diperlukan analisis statistika yang cukup kompleks seperti prosedur analisis faktor. Salah satu prosedur pengujian validitas konstruk yang lebih sederhana adalah dengan pendekatan multi-trait multi-method yang dikembangkan oleh Campbell dan Fiske (Allen dan Yen, 1979). Metode ini digunakan jika dua atau lebih trait yang hendak diukur dengan dua atau lebih metode. Prosedur ini akan menghasilkan adanya bukti validitas diskriminan dan validitas konvergen.

Selain validitas, hal penting lainnya dalam kaitannya alat ukur adalah reliabilitas. Reliabilitas mengacu pada konsistensi pengukuran, yaitu bagaimana skor tes atau hasil penilaian yang lain tetap (tidak berubah, sama) dari satu pengukuran ke pengukuran yang lain. Hasil-hasil penilaian hanya memberikan ukuran unjuk kerja

terbatas yang diperoleh pada waktu tertentu. Kecuali kalau pengukuran dapat menunjukkan layak konsistensi atas kesempatan yang berbeda, penilai yang berbeda, atau sampel yang berbeda dan domain unjuk kerja yang sama. Hasil penilaian yang konsisten sempurna tidak mungkin dapat diperoleh. Banyak faktor yang mempengaruhi hasil penilaian. Jika suatu penilaian dikenakan kepada kelompok yang sama dua kali secara berurutan, beberapa variasi skor dapat terjadi karena adanya fluktuasi pada memori sesaat, perhatian, usaha kelelahan, ketegangan emosional, tebak-tebak dan sejenisnya. Sebaliknya jika dilaksanakan dalam waktu yang lama antara tes pertama dan tes kedua variasi skor kemungkinan disebabkan oleh pengaruh pengalaman belajar, perubahan kesehatan, lupa dan lain-lain. Variasi skor juga mungkin akan terjadi jika tes essay atau penilaian type unjuk kerja siswa lainnya yang dinilai oleh penilai yang berbeda. Variasi skor juga akan terjadi jika digunakan sampel tugas yang berbeda dari domain yang sama.

Linn dan Gronlund (1995), menjelaskan arti reliabilitas sebagaimana yang digunakan dalam tes dan penilaian sebagai berikut:

“(1) reliability refers to the results obtained with an assessment instrument and not to the instrument itself, (2) an estimate of reliability always refers to a particular type of consistency, (3) reliability is a necessary but not sufficient condition for validity, and (4) reliability is primarily statistical”.

Ada beberapa cara untuk mendefinisikan dan menginterpretasikan reliabilitas tes. Sebagai contoh, suatu tes dikatakan reliabel jika skor amatan memiliki korelasi tinggi dengan skor sebenarnya. Kuadrat korelasi antara skor amatan dengan skor sebenarnya dinyatakan sebagai koefisien reliabilitas. Reliabilitas dapat juga dinyatakan sebagai koefisien korelasi antara skor amatan dua tes yang paralel. Jika dua tes yang paralel diberikan kepada kelompok yang sama dan menghasilkan skor amatan yang berkorelasi, korelasi tersebut adalah koefisien reliabilitas. Sayangnya, dalam banyak kasus skor sebenarnya tidak dapat diperoleh, dan adalah sangat sulit untuk membuat dua tes paralel. Oleh karena itu, reliabilitas harus ditaksir dengan metode yang lain.

Allen dan Yen (1979), mengemukakan tiga metode yang umum digunakan untuk menaksir koefisien reliabilitas yaitu: (1) metode tes ulang, (2) metode tes

paralel, dan (3) metode konsistensi internal. Secara umum masing-masing dari ketiga metode tersebut akan menghasilkan taksiran koefisien reliabilitas (r_x), yang berbeda. Jadi yang dihasilkan hanyalah taksiran, karena nilai sebenarnya koefisien ini adalah tidak dapat diamati.

Sesuai dengan namanya, pada metode tes ulang pengambil tes yang sama mengikuti tes dua kali dengan menggunakan tes yang sama kemudian hasilnya dikorelasikan diperoleh taksiran reliabilitas. Metode tes ulang menghasilkan taksiran reliabilitas tes yang sangat beralasan, tetapi metode ini ternyata memiliki beberapa kelemahan. Pertama metode ini sangat potensial terpengaruh oleh *carry-over effect* antar tes, tes pertama sangat mungkin mempengaruhi hasil tes kedua. Kelemahan kedua berkenaan dengan waktu pelaksanaan tes. Interval waktu yang sangat pendek akan membuat *carry-over effect* dalam memori pengambil tes. Sedangkan interval waktu yang lama akan membawa pengaruh pada perubahan informasi.

Taksiran reliabilitas tes paralel adalah korelasi antara nilai amatan dua tes yang paralel. Dalam kenyataannya dua tes yang paralel hanyalah konsep teoritis, sangat sulit untuk membuktikan bahwa dua tes adalah paralel. Oleh karena itu sering digunakan bentuk tes alternatif sebagai pengganti tes paralel. Bentuk tes alternatif adalah setiap dua bentuk tes yang telah disusun dalam rangka untuk membuatnya paralel, dan keduanya mungkin memiliki rerata skor amatan, variansi, dan korelasi dengan pengukuran lain yang sama atau hampir sama. Korelasi antara skor amatan tes pertama dengan skor amatan tes alternatif r_{x2} adalah merupakan taksiran reliabilitas baik untuk tes pertama maupun tes alternatifnya.

Reliabilitas konsistensi internal ditaksir dengan satu kali pelaksanaan tes sehingga permasalahan yang menyertai metode tes ulang dapat dihilangkan. Metode untuk menaksir reliabilitas yang sangat luas telah diketahui dalam pendekatan ini adalah taksiran reliabilitas belah dua. Pada pendekatan ini tes dibagi menjadi dua bagian, yang satu dan lainnya adalah dianggap sebagai tes alternatif, dan pembelahannya diatur sedemikian rupa sehingga keduanya merupakan tes paralel atau pada dasarnya τ -ekuivalen.

Keuntungan utama penaksiran reliabilitas konsistensi internal adalah bahwa hanya diperlukan satu kali tes saja untuk menghitung taksiran reliabilitas. Namun

demikian metode konsistensi internal tidak cocok jika tes tidak dapat dibagi menjadi bagian-bagian yang paralel atau pada dasarnya τ -ekuivalen atau jika tes tidak memiliki butir-butir independen yang dapat dipisahkan.

Menurut Allen dan Yen (1979) ada tiga cara yang biasa digunakan untuk membelah tes menjadi dua bagian yaitu: (1) metode gasal-genap, butir-butir tes dikelompokkan berdasarkan butir-butir bernomor gasal dalam satu kelompok dan butir-butir tes bernomor genap ke dalam kelompok kedua, (2) metode belah dua sesuai dengan nomor urut, jadi misalnya tes terdiri dari 40 butir, maka butir 1 sampai dengan 20 dikelompokkan dalam satu kelompok dan butir nomor 21 sampai dengan 40 ke dalam kelompok lainnya, dan (3) *matched random subset*.

Teknik untuk membelah tes menjadi dua dapat digeneralisasikan untuk membagi tes menjadi lebih dari dua komponen. Sebagai contoh metode gasal genap dapat dimodifikasi untuk membuat tiga komponen dari tes yang terdiri dari 9 butir dapat dikelompokkan menjadi, pertama butir nomor 1, 4 dan 7, kedua nomor 2, 5, dan 8, dan ketiga nomor 3, 6, dan 9. Berdasarkan asumsi tersebut maka pada dasarnya sebuah tes dapat dibagi menjadi N komponen di mana N maksimum adalah sebanyak jumlah butir dalam tes tersebut.

Berdasarkan uraian di atas, maka dalam pembuatan alat ukur dalam dunia pendidikan harus dilakukan secermat mungkin dan disesuaikan dengan kaidah-kaidah yang telah ditentukan oleh ahli-ahli pengukuran di bidang pendidikan. Untuk melihat reliabilitas suatu alat ukur, yang berupa suatu indeks reliabilitas, dapat dilakukan penelaahan secara statistik. Harga ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*).

Untuk mengestimasi harga reliabilitas suatu tes (butir soal berbentuk pilihan ganda (*multiple choice*)) dapat digunakan rumus sebagai berikut .

$$\hat{\alpha} = \frac{R}{R-1} \left(1 - \frac{\sum s_i^2}{s_x^2} \right) \dots\dots\dots(1)$$

dengan :

R : banyaknya butir soal,

s_i^2 : varians butir soal ke-i,

s_x^2 : varians skor total.

Selain hal di atas, hal penting yang harus diingat adalah kualitas soal tes baru dapat diketahui setelah dilakukan analisis soal tes, baik analisis teoritik atau telaah butir atau ada juga yang menyebut dengan analisis kualitatif, dan analisis empirik atau analisis kuantitatif.

1. Analisis Kualitatif (Telaah Butir)

Aspek-aspek yang perlu diperhatikan dalam melakukan telaah butir atau analisis kualitatif adalah aspek materi, konstruksi, dan bahasa/budaya. Soal tes yang digunakan untuk menseleksi calon mahasiswa baru kali ini adalah Tes Potensi Akademik (TPA) yang disusun dengan bentuk objektif atau pilihan ganda. Untuk soal yang berbentuk obyektif pilihan ganda, ada beberapa sub tes yang cara penulisan butirnya menggunakan kaidah-kaidah sebagai berikut.:

1) Aspek materi

- a) Soal sesuai dengan tujuan
- b) Pilihan jawaban homogen dan logis;
- c) Hanya ada satu kunci jawaban yang paling tepat.

2) Aspek konstruksi

- a) Pokok soal dirumuskan dengan singkat, jelas dan tegas;
- b) Rumusan pokok soal dan pilihan jawaban merupakan pertanyaan yang diperlukan
- c) Pokok soal tidak memberi petunjuk ke kunci jawaban;
- d) Pokok soal bebas dari pernyataan yang bersifat negatif ganda;
- e) Gambar, grafik, tabel, diagram, wacana, dan sejenisnya yang terdapat pada soal jelas dan berfungsi;
- f) Panjang pilihan jawaban relatif sama;
- h) Pilihan jawaban yang berbentuk angka atau waktu harus disusun berdasarkan urutan besar kecilnya angka tersebut atau kronologis;

3) Aspek bahasa/budaya

- a) Menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia;
- b) Menggunakan bahasa yang komunikatif;
- c) Tidak menggunakan bahasa yang berlaku setempat (bias budaya);
- d) Pilihan jawaban tidak mengulang kata/kelompok kata yang sama.

2. Analisis Kuantitatif

Pada penelaahan kuantitatif, butir-butir ditelaah berdasarkan data empirik yang diperoleh dari respons atau jawaban pengambil tes. Dengan demikian penelaahan kuantitatif dilakukan setelah tes diadministrasikan atau diujicobakan kepada sejumlah pengambil tes. Penelaahan kuantitatif bertujuan baik untuk menaksir kesalahan pengukuran yang terjadi secara acak maupun untuk mendeteksi ada atau tidak adanya butir-butir tes yang mengandung bias sehingga potensial menyebabkan terjadinya kesalahan sistematis.

Analisis (telaah) kuantitatif dapat dilakukan secara klasik dan secara modern yang menggunakan Item Respon Theory (IRT). Pada penelitian ini, analisis kuantitatif dilakukan berdasar pada teori klasik dan program yang digunakan ITEMAN. Sedangkan telaah dengan menggunakan IRT digunakan program BILOG.

Telaah kuantitatif menurut pendekatan teori tes klasik menghasilkan karakteristik butir yang meliputi tingkat kesukaran (p), daya pembeda (d), dan efektivitas distraktor. Kesesuaian karakteristik butir dengan jenis dan tujuan tes sangat menentukan kualitas butir tes. Pada analisis butir secara klasikal, tingkat kesukaran (p) dapat diperoleh dengan beberapa cara, antara lain: (1) skala kesukaran linier; (2) skala bivariat; (3) indeks Davis; dan (4) proporsi menjawab benar. Cara yang paling mudah dan paling banyak digunakan adalah skala rata-rata atau proporsi menjawab benar atau *proportion correct* (p), yaitu jumlah peserta tes yang menjawab benar pada butir yang dianalisis dibandingkan dengan peserta tes seluruhnya.

Tingkat kesukaran ini (p) mengandung banyak kelemahan, antara lain tingkat kesukaran sebenarnya merupakan ukuran kemudahan butir karena semakin tinggi indeks p maka makin mudah butir tersebut, dan sebaliknya makin rendah p makin sulit. Dengan demikian ini lebih tepat dinamakan indeks atau tingkat kemudahan.

Besarnya tingkat kesukaran berkisar antara nol dan satu. Suatu butir kadang-kadang dikategorikan ke dalam ekstrim sukar yaitu apabila nilai p mendekati nol dan ekstrim mudah apabila nilai p mendekati satu.

Daya pembeda atau daya beda suatu butir tes berfungsi untuk menentukan dapat tidaknya suatu butir tes membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. Tujuan dari penelaahan daya pembeda adalah untuk melihat kemampuan butir tes tertentu dalam membedakan antara pengambil tes yang berkemampuan tinggi dan pengambil tes yang berkemampuan rendah.

Ada beberapa cara yang digunakan untuk menghitung daya pembeda, yaitu: (1) indeks diskriminasi, (2) indeks korelasi, dan (3) indeks keselarasan. Pada penelitian ini hanya dibahas dua cara untuk menghitung daya pembeda dengan metode korelasi yaitu korelasi *point biserial* dan korelasi *biserial*. Korelasi *point biserial* maupun korelasi *biserial* adalah korelasi *product moment* yang diterapkan pada data, variabel-variabel yang dikorelasikan sifatnya masing-masing berbeda satu sama lain. Variabel butir tes bersifat dikotomi yaitu bernilai 1 untuk jawaban benar dan 0 jika jawaban salah. Di sisi lain, variabel skor total atau sub skor total bersifat kontinum yang diperoleh dari jumlah jawaban yang benar. Nilai koefisien korelasi *point biserial* selalu lebih jika dibandingkan dengan nilai koefisien korelasi *biserial*. Koefisien *point biserial* merupakan kombinasi hubungan antara butir tes, kriteria atau skor total, dan tingkat kesukaran. Korelasi *point biserial* cenderung lebih mengutamakan butir tes yang memiliki tingkat kesukaran rata-rata dan akan maksimum apabila tingkat kesukarannya $p = 0.5$ (Hayat, 1997 dan Suryabrata, 1987). Korelasi biserial merupakan korelasi antara butir tes dan kriteria, bebas dari pengaruh tingkat kesukaran butir tes.

Efektivitas distraktor-distraktor dari suatu butir tes ditelaah untuk mengetahui apakah semua distraktor atau pilihan jawaban yang bukan merupakan kunci telah berfungsi sebagaimana mestinya. Artinya apakah distraktor-distraktor tersebut telah dipilih oleh lebih banyak atau semua peserta tes dari kelompok yang berkemampuan rendah sedangkan peserta tes dari kelompok yang berkemampuan tinggi hanya sedikit atau tidak ada yang memilihnya. Efektivitas distraktor dilihat dari dua kriteria, yaitu

(1) distraktor dipilih oleh peserta tes dari kelompok yang berkemampuan rendah, dan (2) pemilih distraktor tersebar relatif proporsional pada masing-masing distraktor (Azwar, 1996).

Dengan demikian jelaslah bahwa untuk menilai kualitas butir tes tidak cukup hanya memperhatikan tingkat kesukaran dan daya pembeda butir tes yang bersangkutan. Penilaian kualitas butir tes juga harus melihat fungsi pilihan jawaban, terutama distraktor-distraktornya, yaitu harus tampak sebagai jawaban yang benar bagi subjek dari kelompok yang berkemampuan rendah. Sebaliknya harus tampak sebagai jawaban yang salah bagi subjek dari kelompok yang berkemampuan tinggi. Sekalipun suatu butir tes terlalu sukar atau terlalu mudah, namun apabila (1) daya pembeda butir tes, dan (2) distribusi jawaban, memenuhi kriteria, maka butir tes tersebut masih dapat diterima sebagai butir tes yang baik. Kriteria yang dimaksud adalah indeks daya pembeda butir tes (r_{bis}) tidak negatif, dan indeks daya pembeda pilihan jawaban negatif kecuali kunci.

Selain secara sendiri-sendiri butir-butir soal tes itu harus baik, maka soal tes secara keseluruhan juga harus valid dan reliabel yang secara lengkap pengertian dan cara menaksir ke dua istilah ini sudah dibahas di muka.

Selain teori pengukuran klasik, dalam pengukuran pendidikan dikembangkan teori pengukuran modern. Teori pengukuran modern dikembangkan karena teori pengukuran klasik memiliki keterbatasan, karena teori pengukuran klasik terdapat keterbatasan karena bersifat *group dependent* dan *item dependent* (Hambleton, Swaminathan dan Rogers, 1991: 2-5).

Group dependent artinya hasil pengukuran tergantung dari kelompok peserta yang mengerjakan tes. Jika tes diujikan kepada kelompok peserta dengan kemampuan tinggi, tingkat kesulitan butir soal akan rendah. Sebaliknya jika tes diujikan kepada kelompok peserta dengan kemampuan rendah, tingkat kesulitan butir soal akan tinggi.

Item dependent artinya hasil pengukuran tergantung dari tes mana yang diujikan. Jika tes yang diujikan mempunyai tingkat kesulitan tinggi, estimasi kemampuan peserta tes akan rendah. Sebaliknya jika tes yang diujikan mempunyai tingkat kesulitan rendah, estimasi kemampuan peserta tes akan tinggi. Kelemahan

pengukuran semacam ini tidak terdapat dalam teori pengukuran modern, yang selanjutnya disebut teori respons butir (*Item Response Theory*).

Teori respons butir memperbaiki keterbatasan yang ada dalam teori pengukuran klasik. Menurut Hambleton, Swaminathan dan Rogers (1991: 2-5) serta Hulin dkk. (1983), teori respons butir bertujuan membentuk : (a) Statistik butir yang tidak tergantung pada kelompok subyek, (b) skor tes yang dapat menggambarkan profisiensi subyek dan tidak tergantung pada taraf kesulitan tes, (c) model tes yang dapat memberikan dasar pencocokan antara butir tes dan level kemampuan, d) model tes yang asumsi-asumsinya mempunyai dukungan kuat, dan e) model tes yang tidak memerlukan asumsi paralel dalam pengujian reliabilitasnya.

Menurut Hambleton, Swaminathan, dan Rogers (1991: 2-5), pemikiran teori respons butir (*Item Response Theory*) didasarkan pada dua buah postulat, yaitu : (a) prestasi subjek pada suatu butir soal dapat diprediksikan dengan seperangkat faktor yang disebut kemampuan (*latent traits*), dan (b) hubungan antara prestasi subjek pada suatu butir dan perangkat kemampuan yang mendasarinya digambarkan oleh fungsi naik monoton yaitu *item characteristic curve (ICC)*. Pernyataan tersebut mengindikasikan bahwa prestasi peserta tes dalam merespons suatu butir tes tergantung dari kemampuan yang dimilikinya. Semakin tinggi kemampuan yang dimiliki semakin baik prestasi yang ditampilkan peserta tes.

Tujuan teori respons butir ialah membentuk parameter butir dan parameter peserta yang bersifat invariants. Tujuan tersebut akan terwujud jika ada kecocokan antara perangkat data tes dengan model yang digunakan. Model yang digunakan akan berlaku atau cocok jika data tes memenuhi asumsi-asumsi dalam teori respons butir.

2). Asumsi-asumsi teori respons butir

Dalam teori respons butir, model matematisnya mempunyai makna bahwa probabilitas subyek untuk menjawab butir dengan benar tergantung pada kemampuan subyek dan karakteristik butir. Ini berarti peserta tes dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan peserta yang mempunyai kemampuan rendah.

Hambleton dan Swaminathan (1985: 16) dan Hambleton, Swaminathan, dan Rogers (1991: 9) menyatakan bahwa ada tiga asumsi yang mendasari teori respon butir, yaitu unidimensi, independensi lokal dan invariansi parameter. Ketiga asumsi dapat dijelaskan sebagai berikut.

Unidimensi, artinya setiap butir tes hanya mengukur satu kemampuan. Contohnya, pada tes prestasi belajar bidang studi matematika, butir-butir yang termuat di dalamnya hanya mengukur kemampuan siswa bidang studi matematika saja, bukan bidang yang lainnya. Pada praktiknya, asumsi unidimensi tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian dan faktor-faktor administratif dalam tes, seperti kecemasan, motivasi, dan tendensi untuk menebak. Memperhatikan hal ini, asumsi unidimensi dapat ditunjukkan hanya jika tes mengandung hanya satu komponen dominan yang mengukur prestasi suatu subyek.

Independensi lokal terjadi jika faktor-faktor yang mempengaruhi prestasi menjadi konstan, maka respons subjek terhadap pasangan butir yang manapun akan independen secara statistik satu sama lain. Asumsi ini akan terpenuhi apabila jawaban peserta terhadap sebuah butir soal tidak mempengaruhi jawaban peserta terhadap terhadap butir soal yang lain. Tes untuk memenuhi asumsi independensi lokal dapat dilakukan dengan membuktikan bahwa peluang dari pola jawaban setiap peserta tes sama dengan hasil kali peluang jawaban peserta tes pada setiap butir soal.

Menurut Hambleton, Swaminathan, dan Rogers (1991: 10), independensi lokal secara matematis dinyatakan sebagai berikut :

$$p(u_1, u_2, \dots, u_n | \theta) = p(u_1 | \theta) \cdot p(u_2 | \theta) \dots p(u_n | \theta)$$

$$= \prod_{i=1}^n p(u_i | \theta) = \prod_{i=1}^n P_i(\theta)^{x_i} \cdot Q_i(\theta)^{1-x_i} \dots \dots (2)$$

keterangan :

i : 1, 2, 3, ...n

n : banyaknya butir tes

$p(u_i | \theta)$: probabilitas peserta tes yang memiliki kemampuan θ yang dipilih secara acak dapat menjawab butir ke- i dengan benar.

Invarian parameter artinya karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal. Kemampuan seseorang tidak akan berubah hanya karena mengerjakan tes yang berbeda tingkat kesulitannya dan parameter butir tes tidak akan berubah hanya karena diujikan pada kelompok peserta tes yang berbeda tingkat kemampuannya.

Menurut Hambleton, Swaminathan, dan Rogers, (1991: 18), invarian parameter kemampuan dapat diselidiki dengan mengajukan dua seperangkat tes atau lebih yang memiliki tingkat kesukaran yang berbeda pada sekelompok peserta tes. Invarians parameter kemampuan akan terbukti jika estimasi kemampuan peserta tes tidak berbeda walaupun tes yang dikerjakan berbeda tingkat kesulitannya. Invarians parameter butir dapat diselidiki dengan mengujikan tes pada kelompok peserta yang berbeda. Invarians parameter butir terbukti jika estimasi parameter butir tidak berbeda walaupun diujikan pada kelompok peserta yang berbeda tingkat kemampuannya.

Dalam teori respons butir, selain asumsi-asumsi yang telah diuraikan sebelumnya, hal penting yang perlu diperhatikan adalah pemilihan model yang tepat. Pemilihan model yang tepat akan mengungkap keadaan yang sesungguhnya dari data tes sebagai hasil pengukuran.

3). Model Logistik

Pada teori respons butir, digunakan model logistik. Ada tiga model logistik dalam teori respon butir, yaitu model logistik satu parameter, model logistik dua parameter, dan model logistik tiga parameter. Perbedaan dari ketiga model tersebut terletak pada banyaknya parameter yang digunakan dalam menggambarkan karakteristik butir dalam model yang digunakan. Parameter-parameter yang digunakan tersebut adalah indeks kesukaran, indeks daya beda butir dan parameter tebakan semu (*pseudoguessing*).

a). Model Logistik Satu Parameter (1P)

Pada model logistik satu parameter, probabilitas peserta tes untuk menjawab benar suatu butir soal ditentukan oleh satu karakteristik butir, yaitu indeks kesukaran

butir. Menurut Hambleton, Swaminathan, dan Rogers (1991: 12), secara matematis model logistik satu parameter dapat dinyatakan sebagai berikut.

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} \text{ , dengan } i : 1,2,3, \dots, n \dots\dots\dots (3)$$

$P_i(\theta)$: probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir i dengan benar

θ : tingkat kemampuan subyek

b_i : indeks kesukaran butir ke- i

e : bilangan natural yang nilainya mendekati 2,718

n : banyaknya butir dalam tes

Parameter b_i merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Misalkan suatu butir tes mempunyai parameter $b_i = 0,3$, artinya diperlukan kemampuan minimal 0,3 pada skala untuk dapat menjawab benar dengan peluang 50%. Semakin besar nilai parameter b_i , maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%. Dengan kata lain, semakin besar nilai parameter b_i , maka makin sulit butir soal tersebut.

b). Model Logistik Dua Parameter (2P)

Pada model logistik dua parameter, probabilitas peserta tes untuk dapat menjawab benar suatu butir soal ditentukan oleh dua karakteristik butir, yaitu indeks kesukaran butir (b_i) dan indeks daya beda butir (a_i). Menurut Hambleton, Swaminathan, dan Rogers (1991: 15), secara matematis model logistik dua parameter dapat dituliskan sebagai berikut.

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \text{ dengan } i : 1,2,3, \dots, n \dots\dots\dots (4)$$

Keterangan :

$P_i(\theta)$: probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir I dengan benar

θ : tingkat kemampuan subjek

- a_i : indeks daya beda dari butir ke-i
- b_i : indeks kesukaran butir ke-i
- e : bilangan natural yang nilainya mendekati 2,718
- n : banyaknya item dalam tes
- D : faktor penskalaan yang dibuat agar fungsi logistik mendekati fungsi ogive normal yang harganya 1,7.

c). Model Logistik Tiga Parameter (3P)

Sesuai dengan namanya, model logistik tiga parameter ditentukan oleh tiga karakteristik butir yaitu indeks kesukaran butir soal, indeks daya beda butir, dan parameter tebakan semu. Dengan adanya tebakan semu pada model logistik tiga parameter, memungkinkan subyek yang memiliki kemampuan rendah mempunyai peluang untuk menjawab butir soal dengan benar. Secara matematis, model logistik tiga parameter dapat dinyatakan sebagai berikut (Hambleton, Swaminathan, dan Rogers, 1991: 17, Hambleton, dan Swaminathan, 1985 : 49).

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \dots\dots\dots (5)$$

Keterangan :

- $P_i(\theta)$: probabilitas peserta tes yang memiliki kemampuan θ dipilih secara acak dapat menjawab butir I dengan benar
- θ : tingkat kemampuan subjek
- a_i : indeks daya beda dari butir ke-i
- b_i : indeks kesukaran butir ke-i
- c_i : indeks tebakan semu butir ke-i
- e : bilangan natural yang nilainya mendekati 2,718
- n : banyaknya item dalam tes
- D : faktor penskalaan yang dibuat agar fungsi logistik mendekati fungsi ogive normal yang harganya 1,7.

Nilai kemampuan peserta (θ) terletak di antara -3 dan +3, sesuai dengan daerah asal distribusi normal. Pernyataan ini merupakan asumsi yang mendasari besar nilai b_i . Secara teoretis, nilai b_i terletak di antara -2 dan $+2$. Suatu butir dikatakan baik jika nilai ini berkisar antara -2 dan $+2$ (Hambleton dan Swaminathan, 1985: 107). Jika nilai b_i mendekati -2 , maka indeks kesukaran butir sangat rendah, sedangkan jika nilai b_i mendekati $+2$ maka indeks kesukaran butir sangat tinggi untuk suatu kelompok peserta tes.

Parameter a_i merupakan daya pembeda yang dimiliki butir ke-i. Pada kurva karakteristik, a_i merupakan kemiringan (*slope*) dari kurva di titik b_i pada skala kemampuan tertentu. Karena merupakan kemiringan, diperoleh semakin besar kemiringannya, maka semakin besar daya pembeda butir tersebut. Secara teoretis, nilai a_i ini terletak antara $-\infty$ dan $+\infty$. Pada pada butir yang baik nilai ini mempunyai hubungan positif dengan performen pada butir dengan kemampuan yang diukur, dan a_i terletak antara 0 dan 2 (Hambleton dan Swaminathan, 1985: 37).

Peluang menjawab benar dengan memberikan jawaban tebakan semu dilambangkan dengan c_i , yang disebut dengan *tebakan semu*. Parameter ini memberikan suatu kemungkinan asimtot bawah yang tidak nol (*nonzero lower asymptote*) pada kurva karakteristik butir (ICC). Parameter ini menggambarkan probabilitas peserta dengan kemampuan rendah menjawab dengan benar pada suatu butir yang mempunyai indeks kesukaran yang tidak sesuai dengan kemampuan peserta tersebut. Besarnya harga c_i diasumsikan lebih kecil daripada nilai yang akan dihasilkan jika peserta tes menebak secara acak jawaban pada suatu butir.

Pada suatu butir tes, nilai c_i ini berkisar antara 0 dan 1. Suatu butir dikatakan baik jika nilai c_i tidak lebih dari $1/k$, dengan k banyaknya pilihan (Hullin, 1983: 36). Jadi misalnya pada suatu perangkat tes pilihan ganda, ada 4 pilihan untuk setiap butir tesnya, butir ini dikatakan baik jika nilai c_i tidak lebih dari 0,25.

4). Fungsi Informasi Butir dan Tes

Fungsi informasi butir (*Item Information Functions*) merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, pemilihan butir tes, dan perbandingan beberapa perangkat tes. Fungsi informasi butir menyatakan kekuatan atau sumbangan butir tes dalam mengungkap latent trait yang diukur dengan tes tersebut. Dengan fungsi informasi butir diketahui butir yang mana yang cocok dengan model sehingga membantu dalam seleksi butir tes. Secara matematis, fungsi informasi butir memenuhi persamaan sebagai berikut.

$$I_i(\theta) = \frac{[p'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \dots\dots\dots (6)$$

keterangan :

i : 1,2,3,...,n

$I_i(\theta)$: fungsi informasi butir ke-i

$P_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

$P'_i(\theta)$: turunan fungsi $P_i(\theta)$ terhadap θ

$Q_i(\theta)$: peluang peserta dengan kemampuan θ menjawab benar butir i

Fungsi informasi butir dinyatakan oleh Birnbaum (Hambleton dan Swaminathan, 1985: 107) dalam persamaan berikut.

$$I_i(\theta) = \frac{2,89a_i^2(1-c_i)}{\left[(c_i + \exp(Da_i(\theta - b_i))) \right] \left[1 + \exp(-Da_i(\theta - b_i)) \right]^2} \dots\dots\dots (7)$$

keterangan :

$I_i(\theta)$: fungsi informasi butir i

θ : tingkat kemampuan subyek

a_i : parameter daya beda dari butir ke-i

b_i : parameter indeks kesukaran butir ke-i

c_i : parameter indeks peluang kebenaran jawaban tebakan semu (*pseudoguessing*) butir ke-i

e : bilangan natural yang nilainya mendekati 2,718

Berdasarkan persamaan fungsi informasi di atas, maka fungsi informasi memenuhi sifat : (1) pada respons butir model logistik, fungsi informasi butir mendekati maksimal ketika nilai b_i mendekati θ . Pada model logistik tiga parameter nilai maksimal dicapai ketika θ terletak sedikit di atas b_i dan nilai tebakan semu butir menurun; (2) fungsi informasi secara keseluruhan meningkat jika parameter daya beda meningkat.

Berdasarkan sifat fungsi informasi butir, untuk mendapatkan informasi butir yang tinggi, perlu dipilih butir yang mempunyai a_i tinggi dan nilai c_i yang rendah tanpa disertai kecocokan nilai b_i terhadap θ_i fungsi informasi butir belum tinggi.

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut (Hambleton dan Swaminathan, 1985: 94). Berhubungan dengan hal ini, fungsi informasi perangkat tes akan tinggi jika butir tes mempunyai fungsi informasi

yang tinggi pula. Fungsi informasi perangkat tes secara matematis dapat dituliskan sebagai berikut.

$$I_i(\theta) = \sum_{i=1}^n I_i(\theta) \dots\dots\dots (8)$$

Nilai-nilai indeks parameter butir dan kemampuan peserta merupakan hasil estimasi. Karena merupakan hasil estimasi, maka kebenarannya bersifat probabilitas dan tidak terlepas dengan kesalahan pengukuran. Dalam teori respon butir, kesalahan penaksiran standar (*Standard Error of Measurement, SEM*) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan *SEM* mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka *SEM* semakin kecil atau sebaliknya (Hambleton, Swaminathan dan Rogers, 1991, 94). Jika nilai fungsi informasi dinyatakan dengan $I_i(\theta)$ dan nilai estimasi *SEM* dinyatakan dengan $SEM(\hat{\theta})$, maka hubungan keduanya, menurut Hambleton, Swaminathan, dan Rogers (1991 : 94) dinyatakan dengan

$$SEM(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \dots\dots\dots (9)$$

D. Model Persamaan Struktural

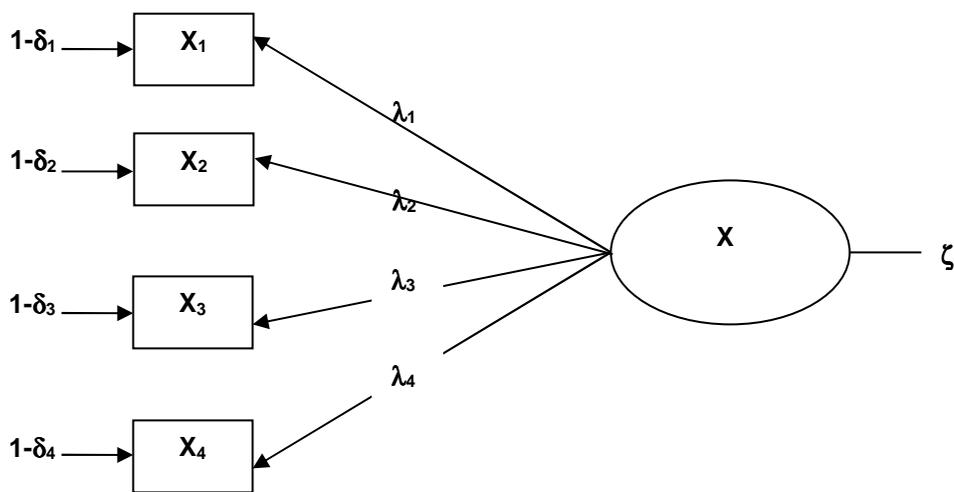
Model Persamaan Struktural merupakan metode statistika yang menggunakan pendekatan uji hipotesis yang terintegrasi antara analisis data dengan konsep konstruksi, untuk melakukan tiga kegiatan secara serempak, yakni memeriksa validitas dan reliabilitas instrumen, pengujian model hubungan antar variabel laten, dan membuat model yang bermanfaat untuk prakiraan. Dengan analisis faktor konfirmatori, validitas dan reliabilitas dari satu alat ukur dapat diestimasi. Validitas dan reliabilitas hasil estimasi selanjutnya dapat dibandingkan berdasarkan nilainya.

Analisis faktor konfirmatori didasarkan pada premis bahwa masing-masing variabel manifest atau variabel yang dapat diamati secara sendiri tidak dapat menggambarkan secara sempurna suatu konsep atau suatu variabel laten atau variabel konstruk. Terkait dengan hal ini, dengan berlandaskan teori, satu konsep atau variabel

laten atau variabel konstruk dapat digambarkan secara bersama oleh beberapa variabel manifes.

Masukan data yang dianalisis pada analisis konfirmatori berupa matriks. Jenis matriks masukan yang biasa dianalisis paling sering berupa matriks kovariansi atau matriks korelasi. Matriks kovariansi dipakai bila variabel-variabel yang dianalisis memiliki skala atau satuan tidak sama, sedangkan matriks korelasi dipakai jika variabel-variabel yang dianalisis memiliki skala atau satuan yang sama (Basuki, 2005). Data ini digunakan untuk mengestimasi faktor strukturnya, dan menggunakan prosedur *maximum likelihood*. Ada dua model analisis konfirmatori, yakni model satu faktor dan model dua faktor. Mengingat pada penelitian ini digunakan analisis konfirmatori model satu faktor, maka selanjutnya hanya akan dibahas analisis konfirmatori model satu faktor.

Misalkan untuk suatu faktor yang memiliki empat indikator ($p = 4$), dapat digambar dalam Gambar 1. Pada Gambar 1, terdapat beberapa parameter, yakni δ (delta), λ (lambda) dan ζ (zeta). Parameter δ (delta) merupakan parameter yang menggambarkan nilai kesalahan pada pengukuran (*measurement error*) pada variabel manifes atau variabel yang diamati (*observed variabel*). Parameter λ (lambda) merupakan parameter yang menggambarkan koefisien struktural (*loading factor*) yang menghubungkan secara linear variabel manifes (variabel yang diamati) dengan variabel laten. Parameter ini berkaitan dengan validitas instrumen, yang dalam hal ini juga diartikan sejauh mana akurasi suatu metode itu mendeteksi DIF. Parameter ζ (zeta) merupakan parameter yang menggambarkan nilai kesalahan pengukuran pada variabel laten berdasarkan variabel manifes atau variabel yang diamati.



Gambar 1. Hubungan antara variabel laten dan variabel indikator pada Analisis Faktor Konfirmatori

Pada persamaan-persamaan pada analisis faktor konfirmatori dengan 4 indikator, persamaan matriks kovariansi untuk parameter model sebagai berikut.

$$\delta_1^2 = \lambda_1^2 + V(\delta_1); \delta_2^2 = \lambda_2^2 + V(\delta_2); \delta_3^2 = \lambda_3^2 + V(\delta_3); \delta_4^2 = \lambda_4^2 + V(\delta_4)$$

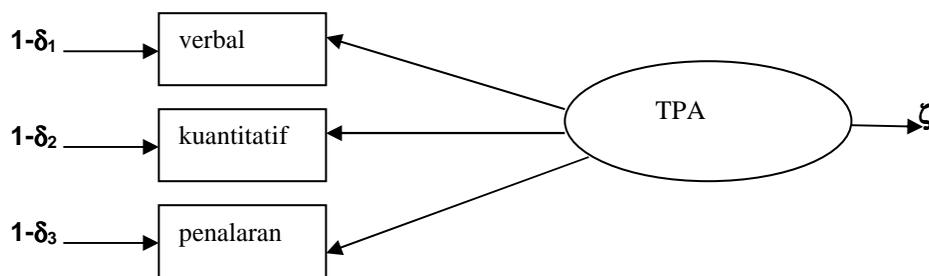
$$\sigma_{12} = \lambda_1\lambda_2; \sigma_{13} = \lambda_1\lambda_3; \sigma_{14} = \lambda_1\lambda_4; \sigma_{23} = \lambda_2\lambda_3; \sigma_{24} = \lambda_2\lambda_4; \sigma_{34} = \lambda_3\lambda_4 \dots (10)$$

Ada 4 model indikator merupakan *overidentified* sebagai 10 persamaan dan 8 parameter yang dapat diestimasi. Persamaan yang *overidentifying* merupakan derajat bebas untuk uji hipotesis, pada kasus model 4 indikator mempunyai 2 derajat kebebasan.

Selain dapat digunakan untuk mengestimasi reliabilitas dan validitas instrumen, analisis konfirmatori juga dapat digunakan untuk mengetahui apakah model hubungan antara variabel-variabel manifes dan variabel laten sudah cocok (fit). Kecocokan ini dapat dilihat dari Khai-kuadrat hitungnya (χ^2). Jika χ^2 lebih besar daripada χ^2 kritis pada tabel, maka dapat dikatakan model yang digambarkan merupakan model yang cocok. Cara lain yang dapat ditempuh juga melihat *p-value*nya. Jika *p-value* lebih besar dari taraf signifikansi (α), maka dapat disimpulkan model yang digambarkan merupakan model yang cocok.

Demikian pula halnya dengan hubungan antar variabel. Informasi hubungan antar variabel biasanya diperlukan dalam penelitian. Hubungan antar variabel-variabel in telah disusun berdasarkan teori tertentu. Dalam model kausal, dikenal variabel eksogen dan variabel endogen, Variabel eksogen merupakan variabel yang keragamannya tidak dipengaruhi oleh penyebab dalam sistem, yang tidak dapat ditetapkan hubungan kausalnya dan memberi efek pada variabel lain (Arief Wibowo, 2003). Sedangkan variabel endogen merupakan variabel eksogen lainnya di dalam model.

Pada penelitian ini, model struktural yang akan diuji dinyatakan pada gambar 2. Kemampuan kuantitatif, kemampuan verbal dan kemampuan penalaran yang dimiliki mahasiswa merupakan variabel indikator yang akan digunakan untuk mengetahui variabel laten, yakni kemampuan TPA. Mengestimasi nilai reliabilitas dan validitas perangkat tes dan menguji kecocokan hubungan-hubungan ini dapat diketahui dengan bantuan program Lisrel.



Gambar 2. Model Analisis Faktor Konfirmatori Tes Seleksi Masuk UNY 2005

E. Kerangka Berfikir

Mengetahui kualitas suatu alat ukur (dalam hal ini seperangkat soal tes) merupakan suatu hal yang sangat diperlukan dalam pelaksanaan tes seleksi masuk. Hal ini dilakukan agar dalam pelaksanaan seleksi, tidak ada pihak-pihak yang dirugikan. Untuk mengetahui kualitas soal tes, dapat dilakukan melalui analisis kualitatif dan analisis kuantitatif. Analisis kualitatif dilakukan sebelum soal itu digunakan, sedangkan analisis kuantitatif dapat dilakukan setelah perangkat soal tersebut digunakan. Analisis kuantitatif dapat dilakukan dengan dua pendekatan, yakni menggunakan pendekatan teori tes klasik maupun teori respons butir. Untuk

mengetahui hubungan antar sub tes, dilakukan dengan menguji kecocokan model hubungan dengan model analisis faktor konfirmatori.

D. Pertanyaan Penelitian

Dengan memperhatikan uraian di atas maka dapat dituliskan bahwa pertanyaan-pertanyaan yang akan dijawab melalui penelitian ini adalah :

1. Bagaimanakah kualitas soal tes masuk UNY program S1 dan D3 Non Reguler dilihat dari materi, konstruksi, dan bahasa ?
2. Bagaimanakah kualitas butir soal tes masuk UNY program S1 dan D3 Non Reguler bila dianalisis dengan pendekatan teori tes klasik?
3. Bagaimanakah kualitas butir soal tes masuk UNY program S1 dan D3 Non Reguler bila dianalisis dengan pendekatan teori respons butir?
4. Bagaimana validitas dan reliabilitas soal tes masuk UNY program S1 dan D3 Non Reguler?
5. Bagaimana hubungan antar subtes yang ada dalam soal tes masuk UNY nonreguler?

BAB III

METODE PENELITIAN

A. Pendekatan Penelitian

Tujuan penelitian ini adalah ingin mengetahui kualitas soal tes masuk UNY program S1 dan D3 Non Reguler 2005. Kualitas tersebut dapat dilihat secara teoritik atau telaah butir, maupun secara kuantitatif atau analisis empiric. Dengan demikian dapat dikatakan bahwa pendekatan yang digunakan dalam penelitian ini, adalah eksploratori.

B. Populasi dan Sampel

Populasi dalam penelitian ini adalah semua butir soal yang ada di soal ujian tertulis yang digunakan dalam seleksi masuk program S1 dan D3 Non Reguler UNY 2005. Sedangkan sampelnya diambil sama dengan populasi, atau dengan kata lain penelitian ini merupakan penelitian populasi, yaitu: semua butir soal yang tersusun dalam seperangkat soal tes masuk UNY program S1 dan D3 Non Reguler 2005 diteliti semua.

Sementara itu, populasi sumber datanya adalah respons atau jawaban peserta tes masuk program S1 dan D3 Non Reguler UNY. Untuk ini semua respons atau jawaban peserta tes masuk UNY program S1 dan D3 Non Reguler 2005 diambil sebagai sampel.

C. Instrumen dan Teknik Pengambilan Data

Dalam penelitian ini, instrumen yang digunakan untuk mengambil data adalah soal tes masuk UNY program S1 dan D3 nonreguler itu sendiri. Secara operasional, responden menjawab pertanyaan-pertanyaan yang ada dalam soal tes dengan cara mengisi lembar jawaban yang telah disiapkan. Selaras instrumen yang telah disiapkan maka teknik pengambilan data dalam penelitian ini adalah tes.

D. Prosedur Penelitian

Prosedur penelitian ini dimulai dari telaah butir, kemudian dianalisis secara empiris baik menurut pendekatan teori tes klasik maupun teori respons butir. Selanjutnya, dilakukan estimasi nilai validitas konstruk dan reliabilitasnya, dan untuk mengetahui hubungan antar subtes, digunakan model persamaan struktural.

E. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah teknik deskriptif kualitatif, deskriptif kuantitatif, analisis butir, dan analisis faktor. Teknik deskriptif kualitatif untuk menjelaskan data hasil analisis butir secara kualitatif, misal kebakuan bahasa yang digunakan dalam butir soal, homogenitas alternatif jawaban, dan lain sebagainya. Teknik deskriptif kuantitatif untuk menjelaskan data kuantitatif, misal jumlah responden yang memilih alternatif jawaban tertentu, jumlah butir yang dikategorikan baik, dan lain sebagainya. Teknik analisis butir empirik digunakan untuk melihat kualitas soal secara empirik, misal: menghitung tingkat kesulitan butir dan daya beda, jika menggunakan pendekatan teori tes klasik dan parameter daya pembeda, tingkat kesulitan dan tebakan semu, fungsi informasi dan kesalahan pengukuran (SEM) jika menggunakan pendekatan teori respons butir. Sedangkan analisis faktor digunakan untuk melihat validitas konstruk soal tes yang sedang dikembangkan.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Ada beberapa hasil yang akan dilaporkan dalam penelitian ini, yaitu: (1) hasil telaah butir, (2) analisis butir empiris dengan pendekatan teori klasik, (3) analisis butir empiris dengan pendekatan teori respons butir, dan (4) hasil estimasi validitas dan reliabilitas perangkat tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler 2005.

A. Telaah Butir

Dalam penelitian ini, analisis butir secara teoretis atau telaah butir dilakukan dengan memperhatikan tiga aspek, yaitu: (1) aspek materi, (2) aspek konstruksi, (3) aspek bahasa/ budaya. Adapun hasil penelaahan pada ketiga aspek itu selengkapnya sebagai berikut.

1. Aspek Materi

Berdasarkan telaah butir yang dilakukan pada soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler diperoleh informasi bahwa: (1) butir-butir dalam paket soal tersebut telah sesuai dengan indikator pencapaian belajar yang diharapkan, (2) distraktor berfungsi sangat baik, dan (3) kunci jawaban untuk tiap-tiap butir soal yang ada hanya satu jawaban. Dengan demikian dapat disimpulkan bahwa dilihat dari aspek materi soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler sudah baik.

2. Aspek Konstruksi

Dari sembilan hal yang harus diperhatikan dalam menelaah butir berdasarkan aspek konstruksi, diperoleh informasi bahwa (1) pokok soal dirumuskan dengan singkat, jelas, dan tegas, (2) rumusan pokok soal dan pilihan jawaban merupakan pertanyaan yang diperlukan, (3) pokok soal tidak memberi petunjuk ke kunci jawaban, (4) pokok soal bebas dari pertanyaan yang bersifat negatif ganda, (5) gambar, grafik, tabel, diagram, wacana dan sejenisnya yang terdapat dalam soal

ditampilkan secara jelas dan berfungsi, (6) panjang pilihan jawaban relatif sama, (7) pilihan jawaban tidak menggunakan pernyataan “semua jawaban di atas salah” atau “semua pilihan jawaban di atas benar” dan sejenisnya, (8) pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya angka tersebut atau kronologis, dan (9) butir-butir tidak tergantung pada jawaban soal sebelumnya. Dengan demikian dapat disimpulkan bahwa berdasarkan hasil telaah butir aspek konstruksi soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler sudah baik.

3. Aspek Bahasa/Budaya

Selanjutnya, apabila dilihat dari aspek bahasa/budaya maka soal itu : (1) menggunakan bahasa yang sesuai dengan kaidah Bahasa Indonesia, (2) menggunakan bahasa yang komunikatif, (3) tidak menggunakan bahasa yang berlaku setempat, dan (4) pilihan jawaban tidak mengulang kata/kelompok kata yang sama.

Berdasarkan hasil telaah dari ketiga aspek yang telah ditentukan, yakni: (1) aspek materi, (2) aspek konstruksi, dan (3) aspek bahasa/budaya maka dapat disimpulkan bahwa butir-butir soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler sudah baik.

B. Analisis Butir Secara Empiris Menggunakan Teori Klasik

Setelah dilakukan telaah butir, langkah selanjutnya adalah analisis empiris menggunakan pendekatan teori klasik. Dalam analisis ini akan ditentukan tingkat kesulitan butir dan daya pembeda butir soal. Untuk mempercepat pengerjaan, analisis dilakukan dengan menggunakan program *BILOG fase I*. Hasil analisis butir secara empiris yang menggunakan pendekatan teori klasik dapat dilihat pada Tabel 1 sampai dengan tabel 3.

Dari Tabel 1 dapat diketahui bahwa ada butir yang kurang baik pada sub tes verbal, yakni butir 1,2,13,14,15,20,26, dan 28. Butir ini kurang baik disebabkan karena daya pembeda yang ditunjukkan dengan korelasi biserial, berharga negatif. Pada tabel 2, yakni pada subtes kuantitatif, hanya ada 1 butir yang kurang baik. Butir yang daya pembedanya negatif pada subtes kuantitatif adalah butir nomor 25. Pada

subtes penalaran, ada 9 butir yang kurang baik, yakni butir nomor 14, 24, 25, 27, 28, 29, 32, 33, dan 35. Butir yang daya pembedanya negatif akan dijawab dengan betul oleh kelompok calon mahasiswa yang kemampuannya di bawah, dan justru akan dijawab salah oleh calon mahasiswa yang kemampuannya tinggi.

Tabel 1. Hasil analisis butir soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler dengan pendekatan teori tes klasik untuk sub tes verbal

No.	Indeks Tingkat kesulitan	Indeks daya pembeda
1	0.316	-0.028
2	0.197	-0.068
3	0.26	0.096
4	0.469	0.133
5	0.662	0.145
6	0.472	0.082
7	0.29	0.054
8	0.166	0.13
9	0.167	0.092
10	0.293	0.078
11	0.274	0.055
12	0.636	0.091
13	0.212	-0.056
14	0.364	-0.128
15	0.145	-0.077
16	0.234	0.064
17	0.476	0.095
18	0.822	0.062
19	0.298	0.073
20	0.091	-0.14
21	0.512	0.125
22	0.179	0.026
23	0.131	0.046
24	0.459	0.159
25	0.333	0.204
26	0.148	-0.123
27	0.179	0.107
28	0.172	-0.056
29	0.64	0.154
30	0.04	0.033

Tabel 2. Hasil analisis butir soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler dengan pendekatan teori tes klasik untuk sub tes kuantitatif

Butir	Indeks Tingkat kesulitan	Indeks daya pembeda
1	0.786	0.305
2	0.934	0.442
3	0.472	0.296
4	0.836	0.387
5	0.398	0.227
6	0.579	0.328
7	0.757	0.445
8	0.829	0.537
9	0.624	0.49
10	0.624	0.388
11	0.172	0.456
12	0.305	0.286
13	0.305	0.518
14	0.46	0.288
15	0.345	0.465
16	0.317	0.488
17	0.484	0.56
18	0.409	0.256
19	0.517	0.356
20	0.497	0.478
21	0.526	0.502
22	0.26	0.437
23	0.553	0.356
24	0.338	0.315
25	0.105	-0.217
26	0.536	0.338
27	0.388	0.342
28	0.183	0.523
29	0.322	0.357
30	0.217	0.256

Tabel 3. Hasil analisis butir soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Regular dengan pendekatan teori tes klasik untuk sub tes penalaran

No.	tk	dp
1	0.697	0.413
2	0.605	0.277
3	0.845	0.413
4	0.736	0.433
5	0.936	0.64
6	0.741	0.022
7	0.302	0.279
8	0.512	0.244
9	0.771	0.435
10	0.512	0.31
11	0.3	0.083
12	0.39	0.257
13	0.748	0.367
14	0.06	-0.29
15	0.464	0.271
16	0.759	0.393
17	0.722	0.477
18	0.741	0.461
19	0.569	0.454
20	0.438	0.234
21	0.817	0.359
22	0.845	0.544
23	0.845	0.463
24	0.033	-0.238
25	0.017	-0.263
26	0.772	0.442
27	0.045	-0.188
28	0.052	-0.097
29	0.155	-0.147
30	0.6	0.359
31	0.591	0.305
32	0.233	-0.002
33	0.057	-0.266
34	0.216	0.257
35	0.064	-0.241

Mencermati butir-butir yang kurang baik berdasarkan teori tes klasik, dapat dijelaskan sebagai berikut. Butir-butir yang ditolak berdasarkan teori tes klasik biasanya tingkat kesulitannya rendah (yang berarti soalnya sulit sekali), atau nilai

korelasi point biserialnya negatif. Kedua hal ini saling mempengaruhi. Jika butir soal terlalu sulit atau terlalu mudah, maka butir soal tersebut variabilitasnya kecil, sehingga tidak mampu mengukur tingkat kemampuan peserta tes. Hal ini mengakibatkan butir soal tersebut daya pembedanya rendah.

Memperbaiki butir-butir yang ditolak berdasarkan teori tes klasik, dapat dilakukan dengan mengurangi faktor kesulitan/kemudahan pada butir soal tersebut. Pada butir soal yang mengandung kelompok kata-kelompok kata, dapat diperbaiki dengan menyusun kembali dengan kelompok kata yang lebih mudah dipahami. Sedangkan untuk kemampuan yang terkait dengan perhitungan, kekompleksan bilangan yang dihitung sebaiknya dipermudah dengan memilih bilangan yang lebih sederhana.

C. Hasil Analisis Butir yang Menggunakan Teori Respons Butir

Butir baik hasil analisis menggunakan pendekatan teori klasik, selanjutnya dianalisis menggunakan teori respons butir. Dengan analisis ini, butir-butir soal dilihat dari tingkat kesulitan (b_j), daya beda (a_j) dan faktor tebakan (c_j) nya. Untuk memudahkan estimasi, analisis dilakukan dengan menggunakan program *Bilog* versi 3,07 fase II.

Dengan program *Bilog* versi 3,07 akan diperoleh tiga kelompok hasil, yaitu: phase 1, phase 2, dan phase 3. Dalam phase 1 dijelaskan statistik butir secara klasik yang meliputi proporsi benar atau tingkat kesulitan butir, korelasi biserial atau daya pembeda butir, dan validitas butir atau korelasi skor butir dengan skor tes. Pembahasan tentang hasil analisis butir secara empiris dengan pendekatan teori klasik telah dipaparkan di atas sehingga tidak perlu diulang lagi.

Dalam phase 2 ditampilkan parameter butir dan probabilitas χ^2 , serta derajat kebebasan untuk masing-masing butir tersebut. Apabila harga χ^2 probabilitas lebih besar dari harga χ^2 kritis maka disimpulkan bahwa butir itu tidak cocok dengan model. Dalam hal ini χ^2 diperoleh dari tabel dengan derajat kebebasan (DF) tiap-tiap butir tadi dan taraf kesalahan yang diambil $\alpha = 0,01$.

Ringkasan hasil estimasi parameter butir pada perangkat tes 51 untuk subtes verbal, serta butir-butir yang cocok dengan model dapat dilihat pada Tabel 4. Dari

tabel 4, dapat diperoleh bahwa pada subtes verbal, ada 5 butir soal yang tidak cocok dengan model logistik 3 parameter, yakni butir nomor 18, 20, 24, 25 dan 30.

Tabel 4. Parameter butir soal 51 untuk subtes verbal dan kecocokannya dengan model logistik 3 parameter

Butir	a	b	c	CHISQ	dk	Keterangan
3	0.412	3.37	0.177	8.7	8	Cocok
4	0.408	1.031	0.194	7.6	6	Cocok
5	0.317	-0.726	0.18	8.4	7	Cocok
6	0.293	1.231	0.178	15.3	8	Cocok
7	0.332	3.732	0.202	5	8	Cocok
8	0.399	4.499	0.122	7.6	7	Cocok
9	0.468	3.973	0.132	2.6	8	Cocok
10	0.403	2.558	0.157	12.1	7	Cocok
11	0.394	2.864	0.155	8.2	8	Cocok
12	0.246	-0.537	0.189	6	8	Cocok
16	0.343	4.576	0.176	6.3	7	Cocok
17	0.309	1.15	0.179	8.7	8	Cocok
18	0.368	-2.241	0.175	25.1	6	Tidak Cocok
19	0.329	4.487	0.234	9.9	8	Cocok
21	0.457	0.489	0.166	20.4	7	Tidak Cocok
22	0.373	5.316	0.147	6.1	7	Cocok
23	0.391	7.716	0.13	6.3	6	Cocok
24	0.457	0.826	0.154	29.3	7	Tidak Cocok
25	0.732	1.17	0.123	30.5	6	Tidak Cocok
27	0.4	4.879	0.146	3.6	7	Cocok
29	0.392	-0.435	0.176	17.7	7	Cocok
30	0.319	*****	0.132	57.5	5	Tidak Konvergen

Pada subtes kemampuan kuantitatif, hanya ada 3 butir yang dianalisis dengan menggunakan teori respons butir merupakan butir yang tidak cocok dengan model 3 parameter, yakni butir 7, 8, dan 9. Hal ini disebabkan karena butir-butir tes kuantitatif merupakan butir yang sedikit sekali menggunakan kalimat-kalimat yang panjang, sehingga meminimasi adanya kemampuan lain yang diperlukan dalam mengerjakan butir-butir soal ini. Hasil analisis butir-butir pada subtes kuantitatif secara lengkap disajikan pada tabel 5.

Pada subtes penalaran, ada 4 butir soal yang tidak cocok dengan model logistik 3 parameter. Butir-butir tersebut adalah butir nomor 19, butir nomor 21, dan

butir nomor 23 dan 35 (abstrak). Hasil analisis butir-butir pada subtes penalaran secara lengkap disajikan pada tabel 7. Mengkaji lebih jauh butir-butir yang tidak cocok dengan model logistik 3 parameter, tidak ada pola yang pasti jika dilihat dari daya pembedanya (a), tingkat kesulitannya (b) ataupun tebakan semunya (c). Ketidakcocokan ini disebabkan karena pola respons siswa terhadap butir ini tidak cocok dengan model 3 parameter, dan perlu penelitian lebih lanjut apakah butir ini cocok dengan model Rasch (1 parameter) atau model 2 parameter.

Tabel 5. Parameter butir soal 51 untuk subtes kuantitatif dan kecocokannya dengan model logistik 3 parameter

Butir	a	b	c	CHISQ	DF	Keterangan
1	0.469	-1.468	0.176	12.6	8	Cocok
2	0.832	-2.276	0.165	8.6	4	Cocok
3	0.454	0.625	0.126	17.8	9	Cocok
4	0.632	-1.595	0.16	6.7	6	Cocok
5	0.469	1.481	0.194	10.3	9	Cocok
6	0.502	0.016	0.158	14.1	9	Cocok
7	0.731	-0.949	0.141	21	7	Tidak Cocok
8	1.186	-1.14	0.121	30.1	5	Tidak Cocok
9	0.841	-0.252	0.128	22.8	7	Tidak Cocok
10	0.602	-0.303	0.135	15.6	8	Cocok
11	1.454	1.707	0.097	11.6	9	Cocok
12	0.983	1.62	0.202	13.6	9	Cocok
13	1.161	1.112	0.13	11.3	9	Cocok
14	0.832	1.179	0.295	16.4	9	Cocok
15	1.297	1.109	0.187	7.3	8	Cocok
16	0.799	1.067	0.091	5.5	9	Cocok
17	1.152	0.465	0.181	17	8	Cocok
18	0.49	1.246	0.177	6.1	9	Cocok
19	0.807	0.769	0.288	8.8	9	Cocok
20	1.014	0.505	0.207	9	8	Cocok
21	1.243	0.486	0.252	10.9	7	Cocok
22	1.086	1.427	0.134	6	9	Cocok
23	0.683	0.518	0.265	8.1	8	Cocok
24	1.434	1.467	0.245	7.1	9	Cocok
26	0.844	0.838	0.33	7.2	8	Cocok
27	0.861	1.369	0.243	4.8	9	Cocok
28	1.349	1.608	0.093	9	9	Cocok
29	0.83	1.53	0.186	5.9	9	Cocok
30	0.679	2.461	0.146	6.4	9	Cocok

Tabel 6. Parameter butir soal 52 untuk sub tes penalaran dan kecocokannya dengan model logistik 3 parameter

Butir	a	b	c	CHISQ	DF	Kecocokan dg Model
1	0.58	-0.834	0.097	9.9	7	Cocok
2	0.324	-0.481	0.102	13.9	9	Cocok
3	0.478	-2.193	0.101	8.9	6	Cocok
4	0.532	-1.188	0.081	12.4	8	Cocok
5	0.855	-2.366	0.091	5.3	4	Cocok
6	0.165	-3.272	0.106	15.2	9	Cocok
7	0.601	1.632	0.127	14.7	9	Cocok
8	0.322	0.359	0.108	4	9	Cocok
9	0.567	-1.33	0.1	7.2	7	Cocok
10	0.387	0.207	0.083	16.3	9	Cocok
11	0.364	2.859	0.167	16.3	9	Cocok
12	0.531	1.183	0.15	13.9	8	Cocok
13	0.433	-1.457	0.103	11.8	7	Cocok
15	0.377	0.598	0.107	8.3	8	Cocok
16	0.556	-1.259	0.108	11.2	7	Cocok
17	0.796	-0.763	0.125	11.6	7	Cocok
18	0.628	-1.077	0.087	8.5	7	Cocok
19	0.918	0.112	0.182	28	7	Tidak Cocok
20	0.331	1.05	0.115	4.6	9	Cocok
21	0.554	-1.745	0.085	17.6	6	Tidak Cocok
22	0.955	-1.431	0.075	7.4	5	Cocok
23	0.754	-1.612	0.089	4.3	6	Cocok
26	0.654	-1.24	0.09	6.9	6	Cocok
30	0.555	-0.323	0.086	5.9	8	Cocok
31	0.451	-0.33	0.082	5.5	8	Cocok
32	0.389	6.448	0.215	5.3	9	Cocok
33	0.312	*****	0.1	22.3	6	Tidak Konvergen
34	0.477	2.364	0.076	17	9	Cocok
35	0.313	*****	0.1	21.7	7	Tidak Konvergen

Mencermati lebih lanjut tabel 4, 5 dan 6 di atas, diperoleh bahwa variasi tingkat kesulitan berbeda-beda. Pada subtes verbal, butir-butir soal penyusunnya tergolong relatif sulit. Hal ini ditunjukkan oleh tingkat kesulitan butir-butirnya yang lebih dari 0, demikian pula halnya dengan sub tes kuantitatif. Namun pada subtes penalaran,

tingkat keculitan butir-butirnya lebih banyak yang negatif, yang berarti butir-butir soal relatif lebih mudah.

D. Validitas dan Reliabilitas Perangkat Tes

Validitas perangkat tes seleksi masuk UNY program S1 dan D3 Non Reguler tahun 2005 dilihat dari validitas konstruksinya. Untuk mengetahui validitas konstruk, digunakan analisis faktor, untuk mengetahui kemampuan-kemampuan apa yang diukur dalam butir-butir soal suatu perangkat tes atau seberapa besar butir soal tersebut bermuatan faktor (*factor loading*). Selain itu dapat diketahui faktor-faktor apa yang diukur oleh sejumlah butir soal (*group factor*) atau seluruh butir soal (*common factor*).

Selain itu, prinsip utama analisis faktor adalah untuk menemukan indeks korelasi antar variabel yang diukur. Ada beberapa asumsi yang harus dipenuhi dari analisis faktor, yaitu : a) komponen matriks di atas 0,2; b) muatan faktor harus mempunyai nilai di atas 0,35 (Garson, 2002, 16); c) korelasi antar variabel dengan pengujian keseluruhan matriks lewat uji Bartlett (*Bartlett test of sphericity*) harus signifikan, dan harga KMOnya 0,60 dianggap dalam kategori cukup (Mueller, 1986 : 54). Hasil ringkasan hasil analisis faktor untuk tiap sub tes disajikan pada Tabel 7.

Tabel 7. Ringkasan hasil analisis faktor untuk sub tesverbal, kuantitatif, penalaran dan bahasa Inggris

No.	Sub Tes	KMO	Uji Bartlett	Komponen Varians	Varians total
1.	Verbal	0,512	Signifikan	9	58,078%
2.	Kuantitatif	0,820	Signifikan	8	47,553%
3.	Penalaran	0,762	Signifikan	10	55,493%

Dari tabel 7 untuk kolom varians total yang dapat dijelaskan, diperoleh bahwa sumbangan perangkat tes 51 untuk mengukur kemampuan verbal calon mahasiswa baru sebesar 58,078% dan untuk mengukur kemampuan kuantitatif sebesar 47,553%. Sedangkan sumbangan perangkat soal 52 untuk mengukur kemampuan penalaran sebesar 55,493%.

Reliabilitas pada teori respons butir disebut dengan fungsi informasi. Dan dari nilai fungsi informasi ini dapat juga diestimasi harga kesalahan baku pengukuran (*standard error of measurement* = SEM). Pada sub tes perangkat tes seleksi masuk UNY program S1 dan D3 Non Reguler, estimasi nilai fungsi informasi dan kesalahan pengukuran disajikan pada tabel 8.

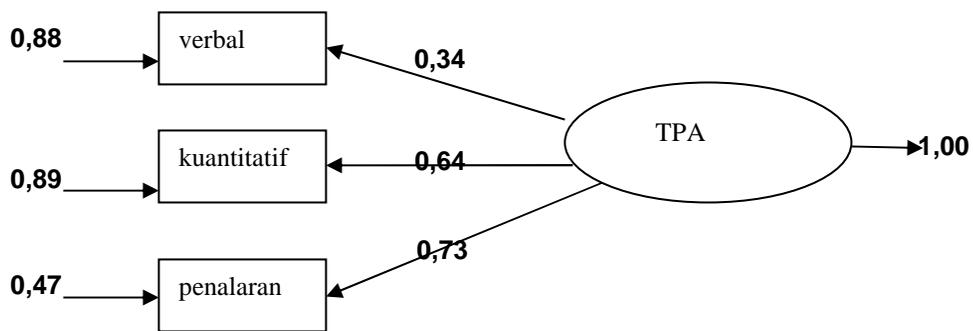
Tabel 8. Hasil estimasi reliabilitas dan kesalahan pengukuran berdasarkan pendekatan teori respons butir

	Subtes	Fungsi Informasi	SEM
1.	Verbal	0,883	1,064
2.	Kuantitatif	9,301	0,327
3.	Penalaran	3,479	0,536

Sebelum mencapai harga maksimumnya, semakin tinggi harga fungsi informasi semakin rendah harga SEM. Sewaktu fungsi informasi mencapai harga tertinggi maka harga SEM mencapai harga terendah. Hal ini sesuai dengan pendapat Jodoin (2003) yang mengatakan bahwa fungsi informasi itu ekuivalen dengan reliabilitas dalam teori tes klasik dan memberikan ukuran kesalahan pengukuran.

E. Hubungan antar Subtes

Hubungan antar subtes, berikut estimasi validitas dan reliabilitasnya hasil keluaran Lisrel disajikan pada gambar 3. Mencermati hasil analisis Lisrel pada gambar 3 tersebut, tampak bahwa model hubungan yang telah digambarkan, dengan variabel latennya tes potensi akademik (TPA) dan variabel indikatornya adalah kemampuan verbal, kemampuan kuantitatif, kemampuan penalaran merupakan model yang cocok (fit). Seperti telah dibahas pada kajian teori, kecocokan model ini tampak dari nilai p-value yang lebih besar dari nilai signifikansi yang ditetapkan, yakni 5%.



p-value = 1,00

Gambar 3. Model Analisis Faktor Konfirmatori Keluaran Lisrel

Memperhatikan nilai λ dan $1 - \delta$ dari masing-masing indikator, dapat diketahui harga-harga hasil estimasinya. Nilai λ merupakan estimasi validitas subtes, sedangkan δ merupakan reliabilitas. Berdasarkan keluaran Lisrel, dapat disarikan tabel estimasi validitas dan reliabilitas tiap subtes.

Tabel 9. Hasil estimasi validitas dan reliabilitas berdasarkan analisis faktor konfirmatori

	Subsub tes	Validitas	Reliabilitas
1.	Verbal	0,34	0,22
2.	Kuantitatif	0,64	0,41
3.	Penalaran	0,73	0,53

Mencermati hasil tersebut, reliabilitas dan validitas subtes verbal pada perangkat tes seleksi masuk UNY Program Non Reguler belum merupakan subtes yang baik. Hal ini terkait juga dengan tingkat kesulitan butir-butir penyusunnya, yang rata-rata merupakan butir yang sulit. Namun reliabilitas dan validitas pada soal penalaran termasuk baik, juga terkait dengan tingkat kesulitan butir-butirnya yang merupakan butir yang relatif mudah.

Dari gambar 3, dapat diperoleh bahwa model hubungan sudah cocok (fit). Dari hasil ini, dapat diketahui bahwa subtes verbal, subtes penalaran dan subtes kuantitatif dapat digunakan sebagai variabel indikator untuk mengukur variabel laten, yakni potensi akademik calon mahasiswa baru. Hal ini sesuai dengan kerangka teori yang digunakan, bahwa Tes Potensi Akademik dapat terdiri dari subtes verbal, kuantitatif, dan subtes penalaran. Dengan kata lain, potensi akademik calon mahasiswa baru nonreguler Universitas Negeri Yogyakarta dapat diukur melalui kemampuan verbal, kemampuan kuantitatif, dan kemampuan penalaran yang dimilikinya.

BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan tujuan, analisis data dan pembahasan yang telah dilakukan maka kesimpulan penelitian ini adalah sebagai berikut.

1. Berdasarkan hasil telaah dari ketiga aspek yang telah ditentukan, yakni: (1) aspek materi, (2) aspek konstruksi, dan (3) aspek bahasa/budaya maka dapat disimpulkan bahwa butir-butir soal tes seleksi masuk UNY tahun 2005 program S1 dan D3 Non Reguler sudah baik.
2. Pada perangkat soal 51, yang terdiri dari subtes verbal, ada 22 butir soal yang baik dari 30 butir soal, pada sub tes kuantitatif, ada 29 butir baik dari 30 butir soal, dan pada perangkat tes 52, yang terdiri dari subtes penalaran, dari 35 butir soal, ada 26 butir soal yang baik, jika ditinjau berdasarkan teori tes klasik.
3. Pada subtes verbal, dari 22 butir soal yang dianalisis, ada 17 butir soal yang cocok dengan model logistik tiga parameter, pada subtes kuantitatif, ada 26 butir dari 29 butir soal yang cocok dengan model logistik 3 parameter, pada subtes penalaran, dari 26 butir yang dianalisis, ada 22 butir yang cocok dengan model logistik tiga.
4. Reliabilitas sub tes verbal, kuantitatif, dan penalaran berdasarkan teori respons butir (nilai fungsi informasinya) berturut-turut adalah 0,883; 9,301; dan 3,479, sedangkan sumbangan sub tes-sub tes tersebut untuk mengetahui kemampuan yang sebenarnya berturut-turut sebesar 58,078%; 47,553%; dan 55,493%.
5. Subtes verbal, subtes kuantitatif dan subtes penalaran merupakan variabel yang dapat digunakan untuk mengestimasi potensi akademik mahasiswa baru Program S1 dan D3 UNY tahun 2005.

B. Saran

Berdasarkan hasil penelitian dan manfaat yang ingin dicapai melalui penelitian ini, maka saran-saran yang diajukan adalah sebagai berikut.

1. Bagi Litbangsisjian UNY, butir-butir soal yang baik beserta karakteristik butirnya perlu didokumentasikan dalam rangka pengembangan bank soal untuk menyusun tes berikutnya.
2. Bagi penyusun perangkat tes seleksi masuk UNY, dalam penyusunan perangkat tes seleksi masuk perlu mempertimbangkan karakteristik butir soal, sehingga perangkat soal mampu mengukur kemampuan calon mahasiswa yang sebenarnya, sehingga perangkat tes bersifat adil, menerima calon mahasiswa yang berhak diterima berdasarkan kemampuannya.

DAFTAR PUSTAKA

- Allen, M.J. & Yen, W.M. 1979. *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ : Prentice Hall.
- Azwar, S. 1996. *Tes prestasi: Fungsi dan pengembangan pengukuran prestasi belajar* (ed. ke 2). Yogyakarta: Pustaka Pelajar.
- Azwar, S. (1997). *Reliabilitas dan validitas* (Edisi 3). Yogyakarta : Pustaka Pelajar.
- Badrun. 2003. Kalibrasi soal ujian masuk UNY program S1 dan D3 sore. *Laporan Penelitian*. Yogyakarta: Tidak diterbitkan.
- Basuki, H. 2005. Pengantar lisrel for windows. *Materi Pelatihan*. Disampaikan pada Pelatihan Pemodelan Persamaan Struktural Angkatan IV tanggal 2-4 Maret 2005 di Lembaga Penelitian Universitas Airlangga Surabaya.
- Brown, F.G. 1974. *Principles of educational and psychological testing* (2nd ed.). New York: Holt, Rinehart and Winston.
- Crocker, L. & Algina, J. 1986. *Introduction to classical and modern test theory*. New York : Holt, Rinehart and Winston Inc.
- Ebel, R.L. 1972. *Essentials of educational measurement*. (3rd. ed.) Englewood Cliffs, NJ: Prentice Hall Inc.
- Fernandes, H.J. X. 1984. *Evaluation of educational program*. Jakarta: National Education Planning , Evaluating and Curriculum Development.
- Garson, D. 2002. *Factor analysis : overview*. Diambil pada tanggal 29 Maret 2003, dari <http://www.class.ncsv/garson/pa765/factoranalysis/>
- Grondlund, N.E. 1982. *Constructing Achievement Test* (3rd. ed.). Englewood Cliffs, NJ: Prentice Hall Inc.
- Guilford, J.P. 1982. *Psychometric Methods* (2nd. Ed.). New York : Tata MacGraw-Hill Publishing Co. Inc.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamentals of item response theory*. London: Sage Publications, Inc.
- Hambleton, R.K. & Swaminathan. 1985. *Item response theory*. Boston, MA: Kluwer Nijhoff, Publisher.
- Jodoin, M.G. 2003. Measurement efficiency of innovative item format in computer-based testing. *Journal of educational*. 40 (1). 1 – 15.

- Linn, R.L. & Gronlund, N.E. 1995. *Measurement and assesment in teaching* (7th ed.). EnglewoodCliffs, NJ. : Prentice-Hall.
- Mehrens, W.A. & Lehmann, I.J. (1973) *Measurement and Evaluation in Education and Psychology*.
New York : Hold, Rinehart and Wiston, Inc.
- Mueller. 1986. *Multivariate analysis*. San Diego, CA : Edits Publisher.
- Nitko, A.J. 1996. *Penilaian berkelanjutan berdasarkan kurikulum (PB2K): Kerangka, konsep, prosedur, dan kebijakan* (terj. AM. Ahmad) Jakarta: Pusat Pengembangan Agribisnis.
- Nunnally, J.C. 1978. *Psychometric theory*. New York: McGraw Hill Book Company. Inc
- Suryabrata, S. 2000. *Pengembangan alat ukur psikologis*. Yogyakarta: Andi Offset.
- Umar, J. 2000. Ujian Akhir sebagai subsistem pendidikan dalam rangka pengendalian mutu. Makalah : disampaikan pada seminar ujian akhir nasional, 29 Agustus 2000, di Ruang Graha Depdiknas Jakarta.
- Wibowo, A. 2005. Pengantar Analisis Jalur. *Materi Pelatihan*. Disampaikan pada Pelatihan Pemodelan Persamaan Struktural Angkatan IV tanggal 2-4 Maret 2005 di Lembaga Penelitian Universitas Airlangga Surabaya.
- Woolfolk, A.E. & McCune, L.N. 1984. *Educational Psychology for Teachers*. Englewood Cliffs, NJ: Prentice Hall, Inc.